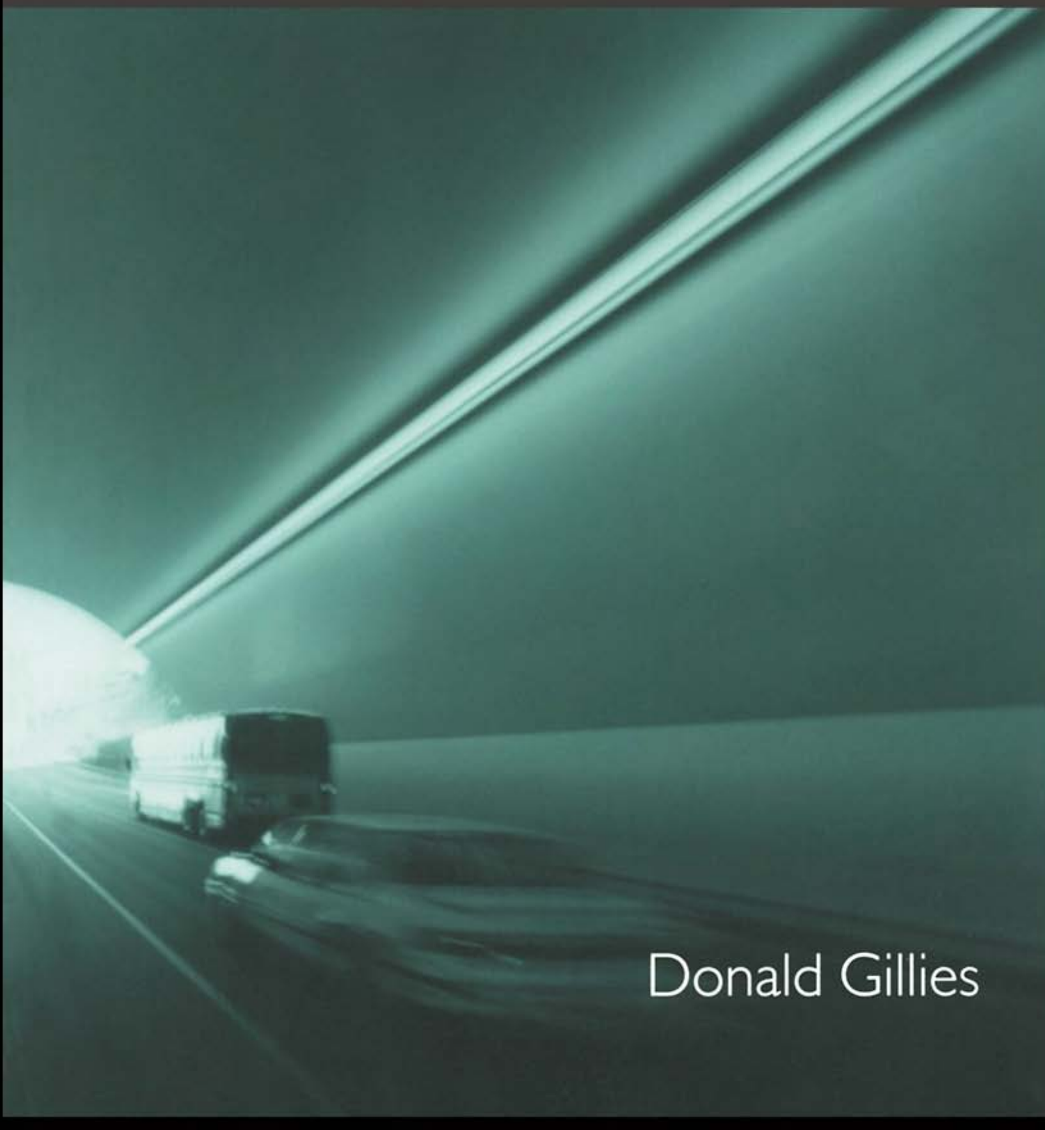
philosophical theories *of* probability



Donald Gillies

Also available as a printed book
see title verso for ISBN details

Philosophical Theories of Probability

'... a very well-presented discussion of the philosophy of probability. A valuable contribution to the field.'

Colin Howson, *London School of Economics and Political Science*

'This book covers ground not adequately covered in any other single text, and does so with great clarity and drive. Its discussion of the case for interpreting probability differently in the social sciences (particularly economics) from in the natural sciences is especially valuable.'

James Logue, *University of Oxford*

The twentieth century has seen a prodigious development of probability and statistics, and their increasing use in almost all fields of research. This has stimulated the creation of many new philosophical ideas about probability. Yet, despite their importance, these ideas tend to be scattered about the literature and not easily accessible.

Philosophical Theories of Probability is the first book to present a clear, comprehensive and systematic account of these various theories and to explain how they are related to one another. It deals with the classical, logical, subjective, frequency and propensity views of probability. The relation of the various interpretations to the Bayesian controversy, which has become central in both statistics and philosophy of science, is explained. Donald Gillies also offers some innovations of his own: a distinctive version of the propensity theory of probability, and the intersubjective interpretation, which develops the subjective theory. He argues for a pluralist view, where there can be more than one valid interpretation of probability, each appropriate in a different context.

This book will prove invaluable to all those interested in the philosophical views of probability and who wish to gain a clearer understanding of the theories and their relations.

Donald Gillies is Professor of Philosophy of Science and Mathematics at King's College, University of London.

Philosophical issues in science
Edited by W. H. Newton-Smith
Balliol College, Oxford

Real History
Martin Bunzl

Brute Science
Hugh LaFollette and Niall Shanks

Verificationism
Cheryl Misak

The Nature of the Disease
Lawrie Reznek

Time, Space and Philosophy
Christopher Ray

Mathematics and Image of Reason
Mary Tiles

Evil or Ill?
Lawrie Reznek

The Ethics of Science: An introduction
David B. Resnik

Philosophy of Mathematics: An introduction to a world of proofs and pictures
James Robert Brown

Theories of Consciousness: An introduction and assessment
William Seager

Psychological Knowledge: A social history and philosophy
Martin Kusch

Is Science Value Free? Values and scientific understanding
Hugh Lacey

Scientific Realism: How science tracks truth
Stathis Psillos

Social Constructivism and the Philosophy of Science
André Kukla

Philosophical Theories of Probability
Donald Gillies

Philosophical Theories of Probability

Donald Gillies

London and New York

First published 2000
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Routledge is an imprint of the Taylor & Francis Group

This edition published in the Taylor & Francis e-Library, 2006.

"To purchase your own copy of this or any of Taylor & Francis or Routledge's
collection of thousands of eBooks please go to www.eBookstore.tandf.co.uk."

British Library Cataloguing in Publication Data
A catalogue record for this book is available
from the British Library

Library of Congress Cataloging in Publication Data
Gillies, Donald, 1944-
Philosophical theories of probability / Donald Gillies.
p. cm. — (Philosophical issues in science)
Includes bibliographical references and index.
1. Probabilities. I. Title. II. Series
BC141 .G55 2000
121'.63—dc21 00–029113

ISBN 0-203-13224-6 Master e-book ISBN

ISBN 0-203-17945-5 (Adobe eReader Format)
ISBN 0-415-18275-1 (hbk)
ISBN 0-415-18276-X (pbk)

To my mother, who first introduced me to philosophy

The need for clarity in scientific and philosophical thought has never appeared to be so essential as today: the most extensive critical analysis of the clearest intuitive concepts can no longer be considered a game for sophists, but is one of the questions which touch most directly on the progress of science.... It is perfectly natural that this need for clarity is felt deeply in the domain of probability, whether because this notion is very interesting from the mathematical point of view as well as from the experimental point of view, or whether because it seems recalcitrant to all attempts to make it precise.

(Bruno De Finetti 1937)

Contents

Illustrations

Figures

Tables

Preface

Probability theory has both a mathematical and a philosophical aspect. The first significant developments in the mathematics of probability took place in the second half of the seventeenth century, and in the same period we find discussions, by among others Leibniz and Locke, of the philosophy of probability. In his *Essay Concerning Human Understanding* of 1690, Locke devotes a chapter (Book IV, xv) to probability. In the preceding chapter he explains his reasons for discussing this subject in the following, somewhat theological, fashion:

> Therefore, as God has set some things in broad daylight, as he has given us some certain knowledge, though limited to a few things ... so, in the greatest part of our concernment, he has afforded us only the twilight, as I may so say, of *probability*, suitable, I presume, to that state of mediocrity and probationership he has been pleased to place us in here ...
>
> (Locke 1690: Book IV, xiv)

This is a most significant passage because it recognises the uncertainty of most of the assumptions which guide our lives, and also looks to the theory of probability as a way of handling this uncertainty.

Since Locke's day the mathematical theory of probability and statistics has developed prodigiously and has come to be used in almost every branch of science. Hand in hand with these mathematical developments have come developments in philosophical ideas about probability. There is now an intricate network of philosophical theories of probability. My aim in this book is to expound these theories as simply and clearly as I can, and to explain how the various views are related to each other.

After some introductory material in Chapter 1, the next seven chapters give systematic expositions of the main philosophical interpretations of probability, which are presented in roughly the historical order in which they were developed. These are the classical, logical, subjective, frequency, propensity and intersubjective views of probability. Some thinkers hold that there is only one correct interpretation of probability, and that the others are mistaken. Such is not the view to be found in the present book, however. In Chapter 8 arguments are put forward for a pluralist conception of probability according to which there is more than one valid interpretation of probability, and different interpretations are suitable for different areas. This last thesis is illustrated in Chapter 9, where it is argued that probability has a different meaning in the natural sciences from its meaning in the social sciences.

Since this book concentrates on the philosophical side of probability, I have tried to make the mathematics used as simple as possible. Probability theory cannot however be understood without some mathematics. The indispensable minimum, which is presupposed throughout, is familiarity with elementary high school algebra, although some knowledge of high school calculus would be useful as well. I do not, however, presume that the reader has studied any probability theory, but rather introduce the basic concepts and axioms, as well as some theorems such as Bayes's theorem, in the course of the book.

Although almost all of the book can be understood with the minimal mathematical knowledge just indicated, there are questions in the philosophy of probability whose formulation and discussion require a more advanced mathematical apparatus. I have dealt with some questions of this sort, but have confined treatment of them to sections marked with a asterisk, e.g. The relation between independence and exchangeability*. In such sections I presuppose familiarity with the standard measure theoretic development of probability theory and with modern mathematical statistics. These sections are arranged so that they can be omitted without losing the main thread of the argument.

Although I hope this book will be of interest to philosophers, particularly philosophers of science, its subject is relevant to many areas outside philosophy. In the epigraph I have chosen for the book, Bruno De Finetti argues that the philosophy of probability is not just 'a game for sophists', but touches 'most directly on the progress of science.' This was an apt observation when it was made in 1937, and it is still more apt today. Indeed, since 1937 probability has entered quite new fields such as econometrics or artificial intelligence, where successful applications do require some consideration of what is the appropriate interpretation of probability. Within statistics the controversy between Bayesians and non-Bayesians continues to rage, and this dispute cannot be properly understood without considering the philosophical aspects of probability. The philosophy of probability lies also at the heart of the mysteries of quantum mechanics. In effect, the subject of this book is important not just for philosophers, but for computer scientists, economists, physicists, statisticians and others as well. The philosophy of probability is one of those theoretical subjects which are also highly relevant for practice.

Donald Gillies
Department of Philosophy
King's College London
June 2000

Acknowledgements

I have been studying and carrying out research in the philosophy of probability for more than thirty years, and during that time I have been helped by contacts and discussions with very many experts in the field. I can perhaps best acknowledge some of these intellectual debts through a brief sketch of the course of my studies.

I was first introduced to the philosophy of probability when, as an undergraduate at Cambridge, I attended Richard Braithwaite's lectures on the subject. He retired at the end of that academic year, and the next year I was able to attend Hugh Mellor's first course of lectures on the subject. Thus while still an undergraduate I had had the benefit of two different points of view on the subject.

I began my PhD in 1966 under the supervision of Imre Lakatos, who at that time was working on a paper on the philosophy of probability (Lakatos 1968) and was still an ardent supporter of Karl Popper's philosophy. Karl Popper had introduced his new propensity theory of probability a few years before. It was therefore a very natural choice to work on this topic for my thesis, and Karl Popper offered a great deal of help and assistance.

It was certainly a great privilege to receive instruction from those two very stimulating philosophers: Imre Lakatos and Karl Popper. Under their guidance, I began as a great enthusiast for objectivism in the philosophy of probability, but subsequent contacts with the subjective school came to alter my viewpoint to some extent. My wife, an Italian economist, had been a student of Bruno De Finetti, and, through her, I met him and had many invaluable discussions concerning the different points of view in the philosophy of probability. I had, moreover, the good fortune to meet and discuss subjective probability with many of the other leading proponents of the school: Phil Dawid, Frank Lad, Dennis Lindley, Marco Mondadori, Jeff Paris, David Spiegelhalter, Peter Williams. In addition, two of my fellow graduate students at the London School of Economics (Colin Howson and Peter Urbach) have subsequently become leading proponents of the subjective Bayesian position (Howson and Urbach 1989). I have a particular debt of gratitude to Colin Howson, with whom I have spent so many hours of friendly discussion over the years on the subjective versus the objective, Bayesian versus non-Bayesian problems. As a result of all this, my own position has shifted from an extreme objectivism to the pluralist view of probability to be found in the present book.

Another intellectual contact which had great importance for me was with the 'Post-Keynesian' school. This group has done most valuable work in studying and reconstructing the development of Keynes's thought. A problem which came to light was the relation between Keynes's early work on the philosophy of probability and the rôle of probability and uncertainty in his later economics. This is connected with the question of the impact of Ramsey's criticisms on Keynes's views on probability. I have had most fruitful discussions on this and related questions with Bradley Bateman, John Davis, Tony Lawson, Jochen Runde and Robert Skidelsky. These interactions suggested the concept of intersubjective probability, which is discussed in Chapter 8, and which is, in a sense, a compromise between the views of Keynes and of Ramsey.

In addition to those already mentioned, many friends and colleagues read sections or the whole of the book and offered valuable comments. I would particularly like to thank in this connection Max Albert, Hasok Chang, David Corfield, Lorraine Daston, Jim Fetzer, Maria Carla Galavotti, Ladislav Kvasz, Moshé Machover, David Miller and Jon Williamson.

Most of the material in the book was presented in the form of lectures for the MSc course in Philosophy and History of Science given jointly by King's College London and the London School of Economics. A lecturer is supposed to teach the audience of students, but this model is something of a myth since the process is in reality one of two-way interaction. Over the years I have received a constant stream of suggestions for improvements and developments from the students attending my lectures. I have been very fortunate to have had such lively and intelligent audiences.

I would like to thank King's College London for allowing me a term of sabbatical leave in the autumn of 1996, and the British Academy for a grant which allowed me a further term of sabbatical leave at the beginning of 1997. It was during this period free from other academic duties that I was able to write the first draft of the present book.

I am grateful to Dover Publications for permission to reproduce material from Richard Von Mises, *Probability, Statistics and Truth,* 2nd Revised English Edition, 1961; to John Wiley & Sons, Inc. for permission to reproduce material from George Soros, *The Alchemy of Finance. Reading the Mind of the Market,* 2nd Edition, 1994; and to Henry Kyburg for permission to reproduce material from his English translation of Bruno De Finetti's 1937 article, which appears in Henry E. Kyburg and Howard E. Smokler (eds.), *Studies in Subjective Probability,* John Wiley & Sons, Inc., 1964, pp. 93–158.

1 Introductory survey of the interpretations

Some historical background

The theory of probability has a mathematical aspect and a foundational or philosophical aspect. There is a remarkable contrast between the two. While an almost complete consensus and agreement exists about the mathematics, there is a wide divergence of opinions about the philosophy. With a few exceptions who will be mentioned later in the book, all probabilists accept the same set of axioms for the mathematical theory, so that they all agree about what are the theorems. Yet in the twentieth century at least, four strikingly different interpretations of this mathematical calculus have been developed, and each of them has adherents today. This book will give a detailed account of these interpretations, but, to orientate the reader, it will, I think, be helpful to begin with an introductory survey of the various views.

Introductory survey of the interpretations

The four principal current interpretations are the following.

1 The *logical* theory identifies probability with degree of rational belief. It is assumed that given the same evidence, all rational human beings will entertain the same degree of belief in a hypothesis or prediction.
2 The *subjective* theory identifies probability with the degree of belief of a particular individual. Here it is no longer assumed that all rational human beings with the same evidence will have the same degree of belief in a hypothesis or prediction. Differences of opinion are allowed.
3 The *frequency* theory defines the probability of an outcome as the limiting frequency with which that outcome appears in a long series of similar events.
4 The *propensity* theory, or at least one of its versions, takes probability to be a propensity inherent in a set of repeatable conditions. To say that the probability of a particular outcome is p is to claim that the repeatable conditions have a propensity such that, if they were to be repeated a large number of times, they would produce a frequency of the outcome close to p.

These four standard interpretations of probability will be described in detail in Chapters 3, 4, 5, 6 and 7. Chapter 8 gives a further interpretation of probability which I introduced in 1991 (see Gillies 1991; Gillies and Ietto-Gillies 1991). This intersubjective view is a development of the subjective theory in which probability is regarded not as the degree of belief of an individual, but as the consensus degree of belief of a social group.

Some advocates of a particular interpretation of probability regard this interpretation as the only valid one. For example, De Finetti, one of the two founders of the subjective theory of probability, thought that all probabilities were subjective in character. By contrast, Popper, who introduced the propensity theory of probability, was not prepared to accept any form of the subjective interpretation. It is, however, possible to argue that one interpretation of probability is valid in one particular context, and another in another. Such pluralist views of probability will be considered in Chapter 8. Perhaps the most famous view of this kind is the two-concept view of probability suggested by Ramsey and developed by Carnap. I will in fact argue for a three-concept view of probability.

Most philosophers of probability agree that the various interpretations of probability can be divided into two broad groups. Unfortunately, there are considerable differences among philosophers about how these two groups should be named. In the next chapter (pp. 18–20) I will discuss these different terminologies, all of which have some advantages and some drawbacks as well. Here I will just give the terminology which I have chosen as on balance the best – though it still has some disadvantages. Interpretations of probability will be divided into (1) *epistemological* (or *epistemic*) and (2) *objective.* The difference is this. Epistemological interpretations of probability take probability to be concerned with the knowledge or belief of human beings. On this approach probability measures degree of knowledge, degree of rational belief, degree of belief, or something of this sort. Clearly the logical, subjective and intersubjective interpretations are all epistemological. Objective interpretations of probability, by contrast, take probability to be a feature of the objective material world, which has nothing to do with human knowledge or belief. Clearly the frequency and propensity interpretations are objective. A favourite example to illustrate this approach is the probability of a particular isotope of uranium disintegrating in a year. Now human beings may know this probability or they may not, but the probability exists quite independently of whether it is known. It exists objectively as a feature of the physical world. Indeed, such isotopes of uranium had this probability of disintegrating in the specified time before there were any human beings at all. To sum up then, epistemological interpretations take probabilities to be related to humans and measures of human knowledge or belief, whereas objective interpretations take probabilities to be human-independent features of the objective material world.

Distinctions are often useful but rarely absolute. In Chapter 8, I will introduce, along with the concept of intersubjective probability, that of *artefactual* probability. It will then be argued that these additional interpretations of probability tend to convert the epistemological/objective distinction into something more resembling a continuum. Despite this slight erosion of the epistemological/objective distinction, it remains, in my view, of fundamental importance for understanding the philosophy of probability, and we will use it constantly throughout this book. I support a pluralist view of probability, and, as an illustration of this position, argue in the concluding chapter that epistemological interpretations are appropriate for economics and the social sciences, whereas the natural sciences require an objective interpretation of probability.

My principal aim in this book is to discuss philosophical views of probability which have been developed during the twentieth century and which are still currently held. Our account of the various interpretations of probability would, however, be incomplete without some mention of the *classical* interpretation of probability expounded by Laplace in his famous *Essai Philosophique sur les Probabilités* (*Philosophical Essay on Probabilities*), first published in 1814. Although there are no advocates of the classical theory today, Laplace's book was enormously influential at the time, and the classical interpretation was the dominant interpretation (or at least very widely held) for at least a hundred years. Some consideration of this theory thus constitutes an essential background to later developments.

Despite the fame of Laplace's *Philosophical Essay on Probabilities,* it is not in fact a very original work. The classical interpretation of probability emerged from discussions in the period roughly from 1650 to 1800, which saw the introduction and development of the mathematical theory of probability. Most of the ideas of the classical theory are to be found in Part IV of Jakob Bernoulli's *Ars Conjectandi,* published in 1713, and Bernoulli had discussed these ideas in correspondence with Leibniz. Nonetheless, it was Laplace's essay which introduced the ideas of the classical interpretation of probability to mathematicians and philosophers in the nineteenth century. This may simply have been because Laplace's essay was written in French and Bernoulli's *Ars Conjectandi* in Latin, a language which was becoming increasingly unreadable by scientists and mathematicians in the nineteenth century.

Because of the historical influence of Laplace's essay, our account of the classical theory in Chapter 2 will be based on Laplace, but in the rest of this chapter we will give a brief account of the historical background to Laplace by sketching some of the main events in the emergence of probability (both mathematical and philosophical) in the period roughly 1650–1800.

Origins and development of probability theory (*c.* 1650 to *c.* 1800): mathematics [1]

The mathematical theory of probability is standardly taken to begin with a correspondence between Pascal and Fermat which took place in 1654. The two mathematicians analysed some gambling problems, the most famous of which had been posed to Pascal by M. le Chevalier de Méré. This is why Poisson was later to say that '*Un problème proposé à un austère janséniste par un homme du monde, a été l'origine du calcul des probabilités.*' (A problem proposed to an austere Jansenist by a man of the world was the origin of the calculus of probability. Quoted from Keynes 1921: v). Here Pascal is the austere Jansenist, and M. le Chevalier de Méré the man of the world. It should be noted, however, that Pascal from the death of his father in 1651 until his religious conversion in his famous *nuit de feu* (night of fire) of 23 November 1654 went through the dissolute period of his life in which he devoted quite a lot of time to gambling.

Of course, no intellectual beginning is ever so abrupt that it can be dated to a particular year. In fact there are a number of predecessors of Pascal and Fermat. Girolamo Cardano (1501–76), an Italian mathematician involved in the solution of the cubic and quartic equations, was a passionate gambler and wrote a treatise *Liber de Ludo Aleae* (*Book of the Game of Dice*). This is mainly a practical handbook for gamblers, but it does contain some mathematical calculations of odds. This treatise was among Cardano's papers at his death but was not published until 1663.

Galileo also devoted a few manuscript pages, probably written between 1613 and 1623, to mathematical problems concerned with dice. Galileo had been consulted by an unnamed gambler who had noticed that, when rolling three dice, 10 is more likely than 9. Yet the number of three-partitions of 10 is the same as that of 9. Galileo solved the problem correctly by an enumeration of possible results. With three dice, there are $6 \times 6 \times 6 = 216$ possible results. Twenty-seven of these give 10, and 25 give 9. So 10 is indeed more probable than 9. (For an English translation of Galileo's text, see David 1962:192–5.) This paper of Galileo's was not published until 1718.

The work of these predecessors did not lead to further developments, whereas, by contrast, the correspondence of Pascal and Fermat marked the beginning of the systematic study and development of the mathematical theory of probability. Huygens, partly inspired by the work of Pascal and Fermat, published his *De Ratiociniis in Aleae Ludo* (*On Calculations in the Game of Dice*) in 1657, and, as David observes: 'This treatise of Huygens ... was, it is said, warmly received by contemporary mathematicians, and for nearly half a century it was the unique introduction to the theory of probability.' (David 1962:115). Huygens' treatise inspired further research into probability, which, from that point on, became a standard field for mathematicians to work in.

It is clear that the stimulus for the introduction of the mathematical theory of probability came from the analysis of gambling games. This is shown by W. Browne's English translation of Huygens' treatise which was published in 1714 with the title: *The value of all chances in games of fortune, cards, dice wagers, lotteries, etc. mathematically determined.* However, this origin of the mathematical theory gives rise to a historical problem, namely 'why was the mathematical theory of probability not developed in the ancient world?' The ancient Greeks were skilled mathematicians, and gambling was very popular in the ancient world. Yet there is no surviving record of any attempt to calculate odds. As Sambursky put it:

> ... we must note with astonishment that, for all the ubiquity and popularity of games of chance, they had no noticeable influence on scientific thought at any time in the Greek and Roman periods. We cannot discover any reference to the formation of the fundamental concepts of probability, ... Nor is there

> any mention of regularities appearing in random series (the law of large numbers), apart from the crudest formulations given by way of illustration.
>
> (Sambursky 1954:179)

This historical problem has been discussed in a number of places. (See, in particular, Hacking 1975:1–10.) I will return to it and suggest a solution in the next chapter. Let us now consider the Pascal–Fermat correspondence in a little more detail.

Pascal's first letter is missing, but the famous problem is contained in his second letter dated Wednesday 29 July 1654. Pascal says:

> I have not time to send you the proof of a difficulty which greatly puzzled M. de Méré, for he is very able, but he is not a geometrician (this, as you know, is a great defect) and he does not even understand that a mathematical line can be divided *ad infinitum* and believes it is made up of a finite number of points, and I have never been able to rid him of this idea. If you could do that, you would make him perfect.
>
> He told me that he had found a fallacy in the theory of numbers, for this reason:
>
> If one undertakes to get a six with one die, the advantage in getting it in 4 throws is as 671 is to 625.
>
> If one undertakes to throw 2 sixes with two dice, there is a disadvantage in undertaking it in 24 throws.
>
> And nevertheless 24 is to 36 (which is the number of pairings of the faces of two dice) as 4 is to 6 (which is the number of faces of one die).
>
> This is what made him so indignant and which made him say to one and all that the propositions were not consistent and that Arithmetic was self-contradictory: ...
>
> (David 1962:235–6)

As we shall see there are good reasons for supposing that, despite his claim at the beginning of the passage quoted, Pascal did *not* have a proof which resolved the difficulty. An analysis of the problem from the modern point of view goes as follows. The chance of *failing* to get a 6 on one throw of a die is 5/6. Therefore on four independent throws it is $(5/6)^4$. Therefore the chance of getting at least one 6 in four such throws is $1 - (5/6)^4 = 671/1296$, or odds of 671:625. Since M. de Méré seems to know the correct value for the odds, it is to be presumed that he had some theoretical method for working it out. He then reasoned, by considering the equality of the ratio of the number of faces and the ratio of the number of throws, that there should be the same chance for getting two 6s in twenty-four throws. However, he had learnt from his gambling experience that the chance in this case was less than rather than greater than 1/2. If now we repeat the above modern argument, we have that the chance of getting two 6s in twenty-four independent throws of two dice is $1 - (35/36)^{24} = 0.4914$ (to four decimal places). Thus, *this* chance is indeed less than a half. David makes the following appropriate comment on the story:

> The Chevalier de Méré was obviously such an assiduous gambler that he could distinguish empirically between a probability of 0.4914 and 0.5, i.e. a difference of 0.0086, comparable with that (0.0108) of the gambler who asked advice of Galileo.
>
> (David 1962:89)

Fermat's reply to this letter from Pascal is unfortunately missing, but we can infer from the subsequent course of the correspondence that he solved the problem correctly. He would not, however, have used the modern method given above, but rather what Pascal later referred to as 'your combinatorial method' (David 1962:239). Fermat's method would have been essentially the same as that used earlier by Galileo on a similar problem. It would have consisted of enumerating the possible results (or combinations) of four throws, and calculating the number of them favourable to getting at least one 6; and similarly in the case of twenty-four throws of two dice. Pascal, however, doubted the validity of this combinatorial method, thereby leading one to suspect that he may not really have had a solution to M. de Méré's problem before receiving Fermat's. In his next letter (Monday 24 August 1654) Pascal writes:

> When there are only *two* players, your combinatorial method is very reliable, but when there are *three,* I think I can prove that it is not applicable unless you proceed in some other way which I have not understood.
>
> (David 1962:239)

Pascal now discusses the following problem. Suppose three men are playing for a stake under the condition that the first to win a certain number of games gets the whole stake. We suppose that the three are equally likely to win any particular game. For some reason it becomes necessary to stop the play when the first man needs *one* game to win, the second needs *two* and the third *two.* How should the stake be divided?

We can solve this problem by Fermat's method of combinations quite easily. The issue would have been settled in at most three games. Let us write a for a win by the first man, b for a win by the second man and c for a win by the third. We have only to write down all the twenty-seven combinations of a, b, c and count the number favourable to the respective men's victories. Thus c c a will be favourable to the third man winning, etc. Following out this method we obtain that the stake should be divided up in the ratio 17:5:5. Pascal starts off correctly but owing to some confused reasoning obtains the result 16:51/2:51/2. Fermat in his reply (Friday 25 September 1654) corrects Pascal's mistake:

> I find only that there are 17 combinations for the first man and 5 for each of the other two: for, when you say that the combination a c c is favourable to the first man and to the third, it appears that you forgot that everything happening after one of the players has won is worth nothing. Now since this combination makes the first man win the first game, of what use is it that the

third man wins the next two games, because even if he won thirty games it would be superfluous?

(David 1962:247–8)

The solution to M. le Chevalier de Méré's problem stimulated mathematicians of the seventeenth century to work out the solutions to a whole sequence of similar gambling problems. We should not perhaps underestimate the economic incentives for this research. Since no methods were known for calculating the correct odds, gambling houses of the time offered on their various games odds which had been determined empirically. Thus, someone who could calculate the correct odds might be able to make considerable gains in a situation in which the empirical odds were somewhat inaccurate.

The next important mathematical advance was taken by Jakob Bernoulli, whose result was published posthumously in 1713 in his *Ars Conjectandi.* Bernoulli proved the first limit theorem concerning probability. We can illustrate his result in modern notation by considering the simple case of tossing a biased coin for which the probability of getting heads [Prob(heads)] is p. It is an elementary result of probability theory that

$$\text{Prob}(r \text{ heads in } n \text{ tosses}) = {}^{n}C_{r}\, p^{r}\, (1 - p)^{n-r} \qquad (1.1)$$ [2]

where ${}^{n}C_{r} = n!/r!\,(n - r)!$ is the number of different ways of choosing r things out of n. This set of probabilities for $r = 0, 1, \ldots, n$ is known as the *binomial distribution.* The simplest limit theorems of probability arise if we consider what happens when n becomes very large. Bernoulli's result is that for any $\varepsilon > 0$,

$$\text{Prob}(\,|\,p - r/n\,| < \varepsilon) \to 1, \text{ as } n \to \infty \qquad (1.2)$$ [3]

Bernoulli's proof gives information about the speed of the convergence, and he illustrates his result with the following example. If $p = 0.6$, and $\varepsilon = {}^{1}/_{50}$, then, if the number of tosses is greater than or equal to 25,550, the odds are 1000:1 that the frequency ratio of heads (r/n) will lie between ${}^{29}/_{50}$ and ${}^{31}/_{50}$.

Bernoulli's theorem is a special case of what have come to be called the *laws of large numbers.* However, there is an ambiguity in this phrase: 'law of large numbers'. We could mean by this a mathematical theorem such as that of Bernoulli just given, or we could mean an empirical law such as the following. If a coin is tossed a large number of times, the observed frequency of heads will tend to a fixed value as the number of tosses increases. Note that this empirical law could be checked by observation without considering any mathematics. But what then is the relation between an empirical law of large numbers and the corresponding mathematical theorem? Does the mathematical result provide a theoretical explanation for the empirical phenomenon? We shall return to this question later on, particularly in Chapter 5, which deals with the frequency theory of probability.

Another limit theorem can be obtained from the binomial Equation 1.1. Here we take the whole of the binomial distribution and consider what happens to it as $n \to \infty$. In fact, this discrete distribution tends to a continuous distribution whose formula is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1.3}$$

This is the famous bell-shaped curve or *normal distribution.* The first to show that the binomial distribution approximated to the normal distribution for large n was De Moivre, who considered only the special case $p = {}^1/_2$ and published the result in 1733. Since then it has been shown that a whole variety of different probability distributions tend to the normal distribution for large n. These results are known as central limit theorems. A graphical illustration of the binomial distribution for p = 0.6 tending to the normal distribution is shown in Figure 1.1.[4]

Another important mathematical result appeared in 1763, when Price published a paper by his friend Bayes after the latter's death. This paper contained Bayes's theorem, and marked the beginning of the Bayesian approach to probability theory. Laplace generalised and improved the results of his predecessors – particularly those of Bernoulli, De Moivre and Bayes. His massive *Théorie analytique des Probabilités,* published in 1812, was the summary of more than a century and a half of mathematical research together with important developments by the author. This book established probability theory as no longer a minority interest but rather a major branch of mathematics.

Origins and development of probability theory (*c.* 1650 to *c.* 1800): practical applications and philosophy

So far we have discussed the analysis of gambling games and mathematical generalisations from these results. However, there were some attempts in this period to apply the mathematical theory to areas other than gambling. These were stimulated by the first attempts to collect and analyse social statistics. A London merchant John Graunt published in 1662 the book *Natural and Political Observations on the Bills of Mortality,* which is an attempt to collect statistics about births and deaths and draw conclusions from them. Some mathematicians tried to apply their theory to empirical material of this kind in a number of cases. There were discussions of the ratio of male to female births, of whether it was worth taking the risk of inoculation against smallpox and most notably the question of life expectancy and the appropriate value for annuities. The inoculation problem was the subject of a controversy between D'Alembert and Daniel Bernoulli (see Daston 1988:82–9). In addition to these statistics-based applications, there were attempts to apply the mathematical calculus to the problem of estimating the probability of someone accused being guilty in the light of the evidence presented, and also to the related problem of the probability of a miracle having actually occurred given testimony that it had occurred.

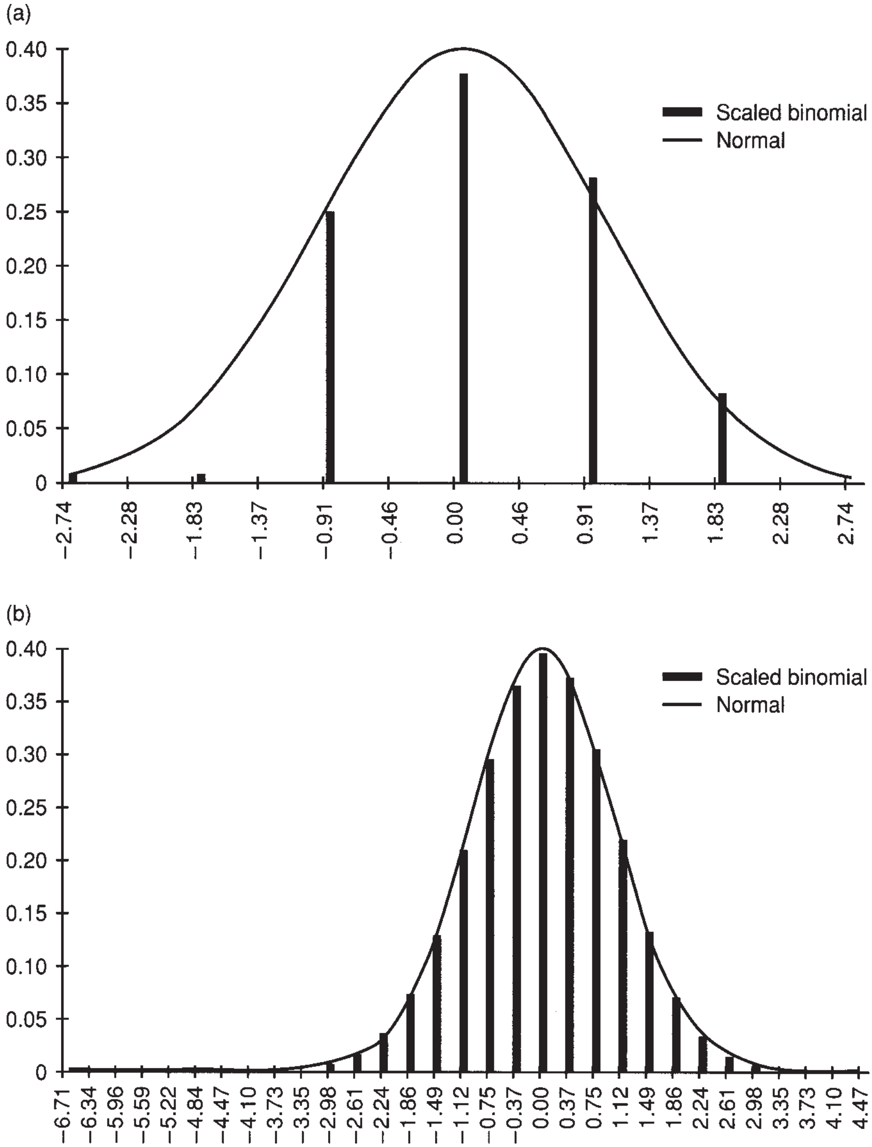


Figure 1.1 How the binomial distribution tends to the normal distribution for increasing n: (a) $p = 0.6$, $n = 5$; (b) $p = 0.6$, $n = 30$.

All these attempted applications are not without interest, but the truth is that they had rather limited success. This is illustrated by the field of annuities. De Moivre published *A Treatise of Annuities upon Lives* in 1725, and yet this seems to have had little impact on the business of annuities in which the rates were estimated from the experience and intuition of the businessman without using mathematical considerations. As Daston observes:

> ... the problems of annuities and life insurance had attracted considerable mathematical attention; yet the terms upon which they were bought and sold had little, if anything, to do with mortality statistics and probability. The actuary was originally a clerical rather than a mathematical position, a combination of secretary and bookkeeper, and with this audience in mind the manuals of De Moivre, Simpson, and Dodson barely required more than arithmetic. Every such book included numerous tables of the values of annuities calculated by age, number of heads, and interest rates to spare the reader calculation. Nonetheless, their impact upon practice appears to have been minimal prior to the establishment of the Equitable Society for the Assurance of Lives in 1762, and even then, the dictates of mathematical theory were greatly tempered by other considerations.
>
> (1988:168–9)

There thus seems to be no escape from the conclusion that, in the period from Fermat and Pascal to Laplace, the principal stimulus for the development of the mathematical theory of probability came from gambling, and the principal practical successes of the theory were in applications to gambling. But this poses a problem. Why did such a serious and important theory which today has so many applications in both the natural and social sciences originate from such a frivolous, and indeed morally dubious, activity as gambling? The answer, I think, is that the standard games of chance involving coins, dice, cards, roulette wheels, etc. can be considered as experimental apparatuses for investigating the phenomenon of randomness. The compulsive gamblers who spent hours studying the outcomes of experimental trials with these pieces of apparatus were in effect scientists conducting a careful examination of the phenomenon of chance and randomness, even though their actual motives were very far removed from those of the disinterested student of nature.

Consider again M. le Chevalier de Méré's discovery that it is a disadvantage to undertake to throw two 6s with two dice in twenty-four throws. His empirical observations at the gambling table had enabled him to realise that a probability of 0.4914 was less than 0.5. The precision of this result is worthy of the finest and most painstaking scientific experimenter, and perhaps M. de Méré should be regarded as such, whatever his actual motives for making these observations.

In the natural sciences, experiments are used to create an artificially simplified environment in which a phenomenon can be studied free from the extraneous and perturbing factors which inevitably occur in real life. The use of experiments enables a scientist to study the phenomenon in this pure state, and hence to ascertain the laws governing it. Once these laws have been mastered, it becomes possible to apply them in the more complicated situations which occur in practice. All this applies exactly to the study of the laws of probability in the context of gambling games, and hence reinforces our analogy between such games and scientific experiments.

It is also noteworthy that the mathematical theory of probability had to be developed for quite a long time in the simplified context of games of chance before it could be successfully applied to practical problems of economic and social importance. Yet when the theory had matured, there was a rich harvest of successful practical applications both in the natural and social sciences, and this possibility was certainly in the minds of many of the early mathematicians who worked in the field. We have seen something of the same kind in the development of artificial intelligence in our own time. Much of the early work was concerned with writing programs to play chess. In itself this was perhaps not very serious. Yet the research led to developments which had more important applications (for some details, see Gillies 1996). Sir Francis Bacon argued that science should be studied for the practical benefits it would yield. Yet at the same time he recognised that there was often the need for a period of theoretical development which would only later yield practical results. He wrote: 'For though it be true that I am principally in pursuit of works and the active department of the sciences, yet I wait for harvest-time, and do not attempt to mow the moss or to reap the green corn.' (Bacon 1620:251).

In the development of many branches of mathematics, one often finds an interplay between problems posed by practical applications and experimental results, purely mathematical developments, and philosophical or foundational discussions. This is certainly true of the development of probability theory in the period from Fermat and Pascal to Laplace, and the same pattern occurs in some very recent developments of probability in artificial intelligence. Let us now turn to some of the philosophical questions discussed in the earlier period.

A few of the philosophical discussions of probability in this period do not appear to have been at all influenced by the mathematical developments just described. An example is Leibniz's first work *De Conditionibus* written in 1665 when Leibniz was nineteen. This is concerned with conditional rights, i.e. rights, such as the right to ownership of a piece of land, which obtain only if a particular condition is fulfilled, e.g. that there is no male heir to the land in the direct line. Leibniz distinguishes three cases. Either the right definitely holds, which is denoted by 1, or it definitely does not hold denoted by 0 or the evidence is not sufficient to determine the case one way or the other, in which case the right is termed uncertain and denoted by a fraction between 0 and 1. These fractions can of course be considered as probabilities. It is noteworthy that Leibniz developed this theory before he had heard of the mathematical work of Fermat, Pascal and Huygens. He only learnt about this during his stay in Paris 1672–6, but naturally he reacted with great enthusiasm. All this goes to show that probability was somehow in the air at that time. Hacking has the following interesting comment on Leibniz's work in this area:

> Leibniz did not contribute to probability mathematics but his conceptualization of it did have lasting impact. Most of his contemporaries started with random phenomena – gaming or mortality – and made some leap of imagination, speculating that the doctrine of chances could be transferred to other cases of inference under uncertainty. Leibniz took numerical probability as a primarily epistemic notion. Degrees of probability

> are degrees of certainty. So he takes the doctrine of chances not to be about physical characteristics of gambling set-ups but about our knowledge of those set-ups. When he went to Paris he found a mathematics tailor-made for his nascent epistemic logic.
>
> (Hacking 1975:89)

This early work of Leibniz is, however, as Hacking implies, something of an exception. Most philosophical discussions of this period are influenced to a greater or lesser extent by the new mathematical calculations of odds in gambling games. An obvious example is Pascal's wager on the existence of God, which appears in his *Pensées* written before his death in 1662 and published in 1670. The relevant passage is no. 418 in Lafuma's numeration and 233 in Brunschvicg's. Here are some extracts from a current English translation:

> Let us then examine this point, and let us say: 'Either God is or he is not.' But to which view shall we be inclined? Reason cannot decide this question. Infinite chaos separates us. At the far end of this infinite distance a coin is being spun which will come down heads or tails. How will you wager? Reason cannot make you choose either, reason cannot prove either wrong....
>
> Yes, but you must wager. There is no choice, you are already committed. Which will you choose then? Let us see: since a choice must be made, let us see which offers you the least interest.... Since you must necessarily choose, your reason is no more affronted by choosing one rather than the other. That is one point cleared up. But your happiness? Let us weigh up the gain and the loss involved in calling heads that God exists. Let us assess the two cases: if you win you win everything, if you lose you lose nothing. Do not hesitate then; wager that he does exist.... here there is an infinity of infinitely happy life to be won, one chance of winning against a finite number of chances of losing, and what you are staking is finite. That leaves no choice; wherever there is infinity, and where there are not infinite chances of losing against that of winning, there is no room for hesitation, you must give everything. And thus, since you are obliged to play, you must be renouncing reason if you hoard your life rather than risk it for an infinite gain, just as likely to occur as a loss amounting to nothing.
>
> (Pascal 1670:150–1)

It is clear that Pascal retained some of the concepts he had acquired as a dissolute gambler after he became an austere Jansenist. Pascal's wager argument was first published in the Port Royal *Logic or the Art of Thinking (La Logique, ou l'Art de Penser)*, written by Pascal's fellow Jansenists (probably Pierre Nicole and Antoine Arnauld). Part IV of this famous and influential book, which was translated into most European languages, was concerned with reasoning under uncertainty and probability.

One form of uncertain reasoning is inductive inference from past evidence to general laws or specific predictions. Concerning such reasoning, Hume posed the famous problem of induction, first in his *Treatise of Human Nature* (1739–40), and then in his *Enquiry concerning the Human Understanding* (originally entitled *Philosophical Essays concerning Human Understanding*) of 1748. As I argue in detail (Gillies 1987), the Bayesianism of Bayes and Price was almost certainly devised as a response to Hume's scepticism about induction. This is an example of a mathematical result (Bayes's theorem of 1763) developed in the attempt to resolve a philosophical problem.

The fourth part of Jakob Bernoulli's *Ars Conjectandi* is concerned with applying the mathematics of probability to civil, moral and economic questions. It is this part of the book which contains Bernoulli's version of the law of large numbers, for this was seen as bridging the gap between the mathematician's probabilities and the social scientist's statistics. This part of the book also contains Bernoulli's philosophical views about which he had corresponded with Leibniz. These views are very much the same as those which Laplace published in his *Philosophical Essay on Probabilities,* and they will accordingly be analysed in the next chapter.

2 The classical theory

The classical theory of probability was a product of the thinking of the European Enlightenment, and it embodied many of the Enlightenment's characteristic ideas. In particular, we find the usual admiration for Newtonian mechanics, and the consequent belief in *universal determinism.* Indeed, Laplace's *Philosophical Essay on Probabilities* of 1814 gives one of the most famous formulations of the thesis of universal determinism. This is the formulation involving what is known as *Laplace's demon.* I will expound it in the next section.

Universal determinism and Laplace's demon

Laplace writes:

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it – an intelligence sufficiently vast to submit these data to analysis – it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past would be present to its eyes.
>
> (1814:4)

The vast intelligence here described has come to be known as Laplace's demon. The idea is obviously founded on that of a human scientist (perhaps Laplace himself) using Newtonian mechanics to calculate the future paths of planets and comets. Extrapolating from this success, it was natural to suppose that a sufficiently vast intelligence could calculate the entire future course of the universe. Laplace himself relates his vast intelligence to human successes in astronomy. As he says:

> The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry, added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world.
>
> (Laplace 1814:4)

Moreover, Laplace explicitly states that the same regularities to be found in the movements of planets and comets exist in all other phenomena as well:

> The regularity which astronomy shows us in the movements of the comets doubtless exists also in all phenomena.
>
> The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; ...
>
> (Laplace 1814:6)

Here we have the typical Enlightenment theme of the superstitions and religious beliefs of former ages being swept away by the triumphant advance of science. Once planets were revered as gods, and comets were regarded with superstitious dread. Now the exact laws governing the movements of planets and comets are understood, and it is possible to predict their future paths with accuracy. In the future this scientific understanding and predictive power will be extended to other phenomena as well. This theme is to be found in many writings of the Enlightenment. One rather elegant expression of it occurs in Gibbon's *Decline and Fall of the Roman Empire.* Gibbon mentions a comet which appeared in 531 AD in the age of Justinian. As he says:

> In the fifth year of his reign, and in the month of September, a comet was seen during twenty days in the western quarter of the heavens, and which shot its rays into the north.... The nations who gazed with astonishment, expected wars and calamities ... and these expectations were abundantly fulfilled.
>
> (Gibbon 1776–88:Vol. V, 249–50)

Gibbon goes on, however, to discuss this comet from the contemporary viewpoint and remarks that 'in the narrow space of history and fable, one and the same comet is already found to have revisited the Earth in *seven* equal revolutions of five hundred and seventy-five years' (1776–88:vol. V, 250). Gibbon describes all seven of these visitations, but we shall confine ourselves to his account of those from the fourth onwards:

> The *fourth* apparition, forty-four years before the birth of Christ, is of all others the most splendid and important. After the death of Caesar, a long-haired star was conspicuous to Rome and to the nations during the games which were exhibited by young Octavian in honour of Venus and his uncle. The vulgar opinion, that it conveyed to heaven the divine soul of the dictator, was cherished and consecrated by the piety of a statesman; while his secret superstition referred the comet to the glory of his own times. The *fifth* visit has been already ascribed to the fifth year of Justinian, which coincides with the five hundred and thirty-first of the Christian era. And it may deserve notice, that in this, as in the preceding instance, the comet was followed, though at a longer interval, by a remarkable paleness of the Sun. The *sixth*

> return, in the year eleven hundred and six, is recorded by the chronicles of Europe and China: and in the first fervour of the Crusades, the Christians and the Mahometans might surmise, with equal reason, that it portended the destruction of the Infidels. The *seventh* phenomenon, of one thousand six hundred and eighty, was presented to the eyes of an enlightened age. The philosophy of Bayle dispelled a prejudice which Milton's muse had so recently adorned, that the comet, 'from its horrid hair shakes pestilence and war.' Its road in the heavens was observed with exquisite skill by Flamsteed and Cassini: and the mathematical science of Bernoulli, Newton, and Halley investigated the laws of its revolutions. At the *eighth* period, in the year two thousand three hundred and fifty-five, their calculations may perhaps be verified by the astronomers of some future capital in the Siberian or American wilderness.
>
> (Gibbon 1776–88:vol. V, 250–1)

Gibbon here contrasts the attitudes to the comet in Roman and medieval times with the way it appeared 'to the eyes of an enlightened age' in which the religious superstitions of former ages had been replaced by the exact science of Newton and Bernoulli. His predictions regarding the next appearance of the comet are also interesting, and they may well prove to be correct.

In the age of Laplace, the successes of Newtonian mechanics inclined most thinkers to accept universal determinism. Developments in the science of our own time, and in particular the development of quantum mechanics, have led to criticisms of universal determinism and a greater inclination towards seeing the universe as indeterministic in nature. It is worth noting that Laplace thought that the same laws would apply to both very large and very small bodies. His vast intelligence 'would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom' (Laplace 1814:4). Daston (1988:244) makes the interesting observation that this inference from macroscopic to microscopic bodies is actually based on the Rules of Reasoning which Newton gave in the *Principia.* Indeed Newton's Rule III runs as follows:

> The qualities of bodies, which admit neither intensification nor remission of degrees, and which are found to belong to all bodies within the reach of our experiments, are to be esteemed the universal qualities of all bodies whatsoever.
>
> (Newton 1687:398)

It follows from this that the laws governing the motion of observable macro-particles (i.e. Newtonian mechanics) should be supposed to hold also for the minute micro-particles of which matter is composed. Quantum mechanics showed that this assumption was false, and that micro-particles such as electrons obey a quite different set of laws. Moreover, these quantum mechanical laws involve probability in an essential way, and hence suggest that the fine structure of the universe may be indeterministic in character. This is a contemporary point of view, but let us now return to the beginning of the nineteenth century and see how Laplace worked out the consequences for probability of his belief in universal determinism.

Equally possible cases

In a completely deterministic system, probabilities cannot be inherent in objective nature but must be relative to human ignorance. Suppose in a particular situation there seem to be three possible outcomes, A, B or C. Because of universal determinism, one of these (A say) *must* occur, and Laplace's demon would be able to foresee the occurrence of A. However, if we humans do not know enough about either the laws of nature or the initial conditions, or both, we may not be able to predict which of A, B or C will occur. In this situation we have to have recourse to the calculus of probabilities. As Laplace puts it:

> The curve described by a simple molecule of air or vapor is regulated in a manner just as certain as the planetary orbits; the only difference between them is that which comes from our ignorance.
>
> Probability is relative, in part to this ignorance, in part to our knowledge. We know that of three or a greater number of events a single one ought to occur; but nothing induces us to believe that one of them will occur rather than the others. In this state of indecision, it is impossible for us to announce their occurrence with certainty.
>
> (Laplace 1814:6)

In this situation where 'nothing induces us to believe that one of them will occur rather than the others', we must, Laplace thinks, regard the events as equally possible. Moreover, the calculus of probabilities can only be applied where we have a number of equally possible cases. Suppose there are *n* such cases, and *m* of them are favourable to the outcome A. Then the probability of A [Prob(A)] is defined to be

$$\text{Prob(A)} = m/n$$

This is the famous classical definition of probability based on equally possible cases. From this definition the standard axioms of probability follow immediately (at least with *finite additivity*).[1] A simple example of the classical definition is afforded by a regular die, for which we have six equally possible cases 1, 2, ..., 6. Of these three (1, 3, 5) are favourable to the outcome 'Odd', whose probability is thus $^3/_6 = {}^1/_2$.

Laplace himself gives the definition as follows:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining

> the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favourable cases and whose denominator is the number of all the cases possible.
>
> (Laplace 1814:6–7)

This approach to probability was the dominant one among mathematicians for nearly a century. For example, the Russian mathematician Markov in 1912 published a book on probability containing many remarkable mathematical developments, such as the theory of Markov chains. However, he still adopts the classical definition as the foundation for the calculus. This wide acceptance for such a long period is in some ways quite surprising for there are seemingly rather obvious objections to the classical theory which was forcibly stated by Von Mises.

Von Mises asks the question: 'But how are we to deal with the problem of a biased die by means of a theory which knows only probability based on a number of equally likely results?' (1928:69). Indeed, there does not seem to be any way in which the classical theory can deal with the case of a biased die, and yet surely one does not want to exclude the case of a biased die from the theory of probability.

Laplace does mention the case of a biased coin in Chapter VII of his *Philosophical Essay,* which is entitled 'Concerning the Unknown Inequalities which may exist among Chances which are supposed to be Equal' (1814:56). Moreover, in his mathematical work on probability he considers a case in which the chance of heads is say $(1 + \alpha)/2$ and of tails is $(1 - \alpha)/2$, and proceeds to do calculations with these quantities (cf. Todhunter 1865:470, 598). This seems to imply the existence of an objective, and possibly unknown, probability of getting heads with a particular coin and so to contradict Laplace's own view that probability is just a measure of human ignorance. It looks as if Laplace forgot his philosophical foundations when developing the mathematical theory.

Janus-faced probability

A more sympathetic attitude would be that Laplace's confusion arises form his partial, but not complete, recognition of what Hacking has recently called the Janus-faced character of probability. Janus was the Roman god who gave his name to our month of January. He was the god of beginnings and was represented with two faces, perhaps one looking back to the past and the other forward to the future. Hacking has argued that

> ... probability ... is Janus-faced. On the one side it is statistical, concerning itself with stochastic laws of chance processes. On the other side it is epistemological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background.
>
> (Hacking 1975:12)

Daston claims that the distinction between the two faces of probability was first made by Poisson in 1837, and Cournot and Ellis in the early 1840s. As she says:

> Since the 1840s, when Cournot and Robert Ellis reassessed the foundations of mathematical probability in terms of relative frequencies, the distinction between, in Cournot's words, "objective possibility," which denotes "the existence of a relation which subsists between things themselves," and "subjective probability," which concerns "our manner of judging or feeling, varying from one individual to the next," has been the departure point for all discussions concerning the interpretation of probability theory. For the greater part of the eighteenth century, however, probabilists would have found such a distinction alien. Their work accommodated both objective and subjective senses of probability with an ease that has bemused later commentators.
>
> (Daston 1988:191)

Daston is undoubtedly right that a dual classification of interpretations of probability has been a feature of discussions of the foundations of the subject since the 1840s. There has, however, as we have already remarked, been considerable disagreement among probabilists concerning the terminology used to mark the distinction. Let us now review some of the suggestions which have been made.

Daston's terminology is employed by Popper, who writes:

> A *subjective interpretation* of probability theory ... treats the degree of probability as a measure of the feelings of certainty or uncertainty, of belief or doubt, which may be aroused in us by certain assertions or conjectures. In connection with some non-numerical statements, the word 'probable' may be quite satisfactorily translated in this way; but an interpretation along these lines does not seem to me very satisfactory for numerical probability statements ... the *objective interpretation,* treats every numerical probability statement as a statement about the *relative frequency* with which an event of a certain kind occurs within a *sequence of occurrences.*
>
> (Popper 1934:148–9)

It is interesting to note that Popper here identifies the objective interpretation of probability with the frequency theory, although later (in 1957) he would introduce a new objective interpretation of probability – the propensity theory. I will discuss this change in detail in Chapter 6. Characteristically, Popper also rejects the subjective approach to probability. This was a constant feature of his thinking on the subject throughout his life.

The difficulty with this terminology is that the 'subjective' interpretations of probability include both the subjective theory of probability, which identifies probability with degree of belief, and the logical theory, which identifies probability with degree of rational belief. Thus, subjective is used both as a general classifier and for a specific theory. This is surely unsatisfactory. The same objection can be made against the terminology I used in my earlier book where I named the two faces of probability the *logical* and the *scientific* (see Gillies 1973:1). Here it is the word 'logical' which is used both as a general classifier and for a particular interpretation. Hacking showed a way out of these difficulties by suggesting that this group of theories could be called *epistemological* or *epistemic.* This seems to me an excellent suggestion since the term 'epistemological' can conveniently be used to cover theories which identify probability with degree of knowledge or ignorance, or degree of belief or degree of rational belief. I have therefore adopted this terminology in the present book.

Let us now turn to the other face of probability. For this Hacking suggests the term *aleatory.* This I find unsatisfactory for the perhaps frivolous reason that this word of Latin extraction is rather difficult to pronounce in English. In my earlier book on probability, I used the term 'scientific'; the idea being that this was the type of probability which appeared in the theories of natural science such as physics or biology. However, my subsequent research has exposed a difficulty in this terminology. I will argue in Chapter 9 that the type of probability which is appropriate for economic theories is epistemological. However, I do not want to imply by this that economic theories are necessarily unscientific. Thus I finally decided for Popper's term, objective, and so classify interpretations of probability as epistemological or objective.

It is only fair to point out that, although this terminology seems to me on balance the best, it does involve a difficulty. Some versions of the logical theory, including that of Keynes which we shall consider in the next chapter, regard probability relations as existing in a kind of Platonic world whose contents can be intuited by the human mind. Thus, this kind of theory, though epistemological, takes probabilities to be in some sense objective. To overcome this difficulty, I suggest distinguishing 'objective in the material sense', meaning referring to objects in the material world, from 'objective in the Platonic sense', meaning referring to objects in a hypothetical Platonic world. When we use 'objective' on its own, it is to be understood as 'objective in the material world'. The other sense of objective will always be qualified as 'objective in the Platonic sense'. So when we classify theories of probability as epistemological or objective, objective is to be understood as referring to objects in the material world. Any choice of terminology reflects a particular theoretical perspective, and my choice here is connected with my disbelief in the existence of a Platonic world of abstract objects.

The analysis of the notion of objective is a difficult matter. Indeed, one could say that it is one of the most fundamental problems in philosophy. In Chapter 8 I will pursue the question further by introducing the concepts of 'intersubjective' and 'artefactual'. For the moment, however, let us leave these questions of terminology, and, returning to the classical theory of probability, ask whether it should be classified as epistemological or objective.

On this point there is room for some debate. Hacking writes:

> In short, around 1660 a lot of people independently hit on the basic probability ideas. It took some time to draw these events together, but they all happened

> concurrently.... It is notable that the probability that emerged so suddenly is Janus-faced.
>
> (1975:11–12)

This seems to imply that in the early period up to Laplace, probabilists had already distinguished the epistemological and objective faces of probability. Yet Daston, as we have seen, argues that, until well into the nineteenth century, no distinction was made between the epistemological and objective senses of probability. Probabilists before that, she claims, 'would have found such a distinction alien' (Daston 1988:192). She also proposes an ingenious theory why the distinction was not made. In her view the general acceptance of an associationist psychology made the distinction superfluous. As she says:

> Experience generated belief and probability by the repeated correlation of sensations which the mind reproduced in associations of ideas. The more constant and frequent the observed correlation, the stronger the mental association, which in turn intensified probability and belief. Hence, the objective probabilities of experience and the subjective probabilities of belief were, in a well-ordered mind, mirror images of one another. This was why intuitive judgments based on broad experience could be trusted. If classical probabilists took the reasonable man as their standard, it was partially because his reasonableness was intrinsically probabilistic.
>
> (Daston 1988:197)

My own view is that probabilists of the period up to Laplace regarded probability as epistemological rather than objective. It is true, as we have seen, that Laplace occasionally uses phrases which seem to imply the existence of unknown chances, but this I would interpret as a slip or inconsistency rather than a commitment to objective probability. Since all probabilists of that period had a firm belief in universal determinism, it is difficult to see how they could have conceived of probability other than as a measure of human ignorance.

This conclusion is reinforced by Laplace's attitude to an interesting example which he gives. This example is valuable, because it illustrates in a striking fashion the difference between epistemological and objective probability. Laplace supposes that someone (Ms A say) is reliably informed that a coin is biased but not told the direction of the bias, and that she is asked to say what is the probability of heads. If Ms A holds an epistemological view of probability, she will answer that Prob(heads) = 1/2, since, because of her ignorance of the direction of the bias, there is no reason to prefer one outcome to the other. If Ms A holds an objective view of probability, she will answer that Prob(heads) = p where $0 = p = 1$, and the value of p is otherwise unknown except for the fact that $p \neq {}^1/_2$. This shows in a striking fashion the difference between the two approaches to probability. According to one, Prob(heads) is exactly one-half. According to the other, all we know about Prob(heads) is that it does not equal one-half. Laplace himself comes out unequivocally in favour of the epistemological view. He writes:

> But if there exist in the coin an inequality which causes one of the faces to appear rather than the other without knowing which side is favored by this inequality, the probability of throwing heads at the first throw will always be 1/2; because of our ignorance of which face is favored by the inequality the probability of the simple event is increased if this inequality is favorable to it, just so much as it is diminished if the inequality is contrary to it.
>
> (Laplace 1814:56)

I will conclude this account of Laplace's ideas on probability by giving one of his famous sayings which seems to me to have particular relevance today. It runs as follows: '... the theory of probabilities is at bottom only common sense reduced to calculus; ...' (Laplace 1814:196). In the modern context one could say that artificial intelligence is at bottom only common sense reduced to calculus, for in artificial intelligence one considers an intelligent human action which the practitioners carry out using their educated common sense, e.g. medical diagnosis, and tries to produce a mathematical model of the procedure which will enable it to be carried out by a computer. I will conclude the present chapter with a section dealing with a historical problem mentioned in Chapter 1.

Why was probability theory not developed in the Ancient World?

No doubt there were many factors which prevented the development of probability theory by the ancient Greeks. However, our analysis of the origins of the theory in the seventeenth century does suggest two things which may have been important. The ancient Greeks were very skilful mathematically, but the principal area of their expertise was geometry. Now the development of probability theory required arithmetic and algebra – precisely the areas which the Greeks tended to neglect. The Greeks had a poor system for representing numbers and for carrying our arithmetical computations, whereas the mathematicians of the seventeenth century had our modern Indian/Arabic decimal system. As for algebra, the Greeks had only a cumbersome geometrical algebra, and our modern system of elementary algebra was developed in the century or so before 1650. Indeed Cardano, who made some of the earliest probability calculations, was also involved in attempts to solve the cubic equation. Could the binomial distribution have been formulated without a good algebraic notation? Could the limit theorems of Jakob Bernoulli and De Moivre have appeared without developments in both algebra and calculus? The Greeks were both ardent gamblers and skilled mathematicians, but their mathematics was just not suitable for analysing gambling.

There was moreover another factor which would have militated against the development of probability theory in the ancient world. We have seen that the first probability problems to be solved concerned regular dice. The assumption that all the faces were equally possible was crucial for the method of solution, which consisted of counting all the possible outcomes and dividing this into the number of outcomes favourable to the result sought. This method simply could not have been applied to irregular dice. Now gambling in the ancient world was mainly carried out not with dice in the modern sense but with *astralagi.* An astralagus is a small bone found in the heels of sheep or deer. It has four flat sides, on which it can rest, and two rounded sides. The flat sides consist of two pairs of opposite sides which were numbered 1, 6, and 3, 4. The numbers 2 and 5 were omitted. One was known as the dog and was considered to be bad. The best result was to get different numbers on throwing four astralagi. This was known as the 'Venus'. David found from rolling a modern sheep's astralagus that 'the empirical probabilities were approximately one in ten each of throwing a 1 or of throwing a 6 and about four in ten each, of throwing a 3 or a 4' (1962:7). Moreover, these empirical probabilities presumably varied from astralagus to astralagus. It would have been very difficult to make a start at calculating odds in such a complicated situation.

There were also dice in our sense in the ancient world, but most of these were irregular, and the few regular ones would not have been widely used in gambling games. As David says:

> The classical dice vary considerably in the materials of which they are made and in the care with which they have been fashioned. The impression left by many of them is that the maker picked up any convenient piece of stone, or wood, or bone and roughly shaped, marked and used it.... This is possibly not surprising since even these imperfect dice must have seemed good after the astralagi. There are exceptions. A few of the dice I have seen are beautifully made, with tooled edges, and throw absolutely true.
>
> (1962:10–11)

It might be objected[2] that after all the Greeks and Romans had well-balanced coins, and that these could have been used to make a start with probability theory. However, in the seventeenth century nearly all the early probability calculations are concerned with dice. This was no doubt because the important gambling games were with dice. In the ancient world they would have been with astralagi.

If it is indeed correct that the irregularities of astralagi inhibited the development of probability theory, this provides a kind of historical justification for the classical theory of probability. This theory bases probability on equally possible cases, and, indeed, in the early days of probability theory mathematical calculation were only possible under this simplifying assumption. As long as probability theory dealt mainly with regular coins and dice, and well-shuffled packs of cards, the classical theory of probability did indeed provide an adequate foundation for the subject. From the middle of the nineteenth century onwards, however, probability theory came to be applied more and more in the natural sciences (physics and biology) and in the social sciences and economics. The old foundation in terms of equally possible cases was no longer adequate for these new applications, and so, throughout the twentieth century, there has been a sequence of attempts to provide a better foundation for the subject. The recent novel applications of probability in artificial intelligence provide an important current stimulus for continuing this work.

Yet it is not just the whole range of modern applications of probability which render the classical theory of probability unsatisfactory for us. The classical theory embodies assumptions which were held by most thinkers in the age of the enlightenment but which no longer seem plausible to us today. One such assumption is universal determinism, but another, emphasised by Daston in her 1988 book, is that of the reasonable man whose reasoning based on experience is an accurate reflection of what goes on in the world, so that no sharp distinction need be made between subjective and objective. The eighteenth century was after all the age of reason, but the same could not be said of the twentieth century.

What has been characteristic of the twentieth century is on the one hand the use of a mathematical and scientific apparatus far exceeding in power anything that existed in the eighteenth century, but on the other the prevalence of outbursts of violence and of beliefs with no rational or scientific foundation. An obvious case of this contradiction is Hitler's Germany, where the country's skills in mathematics and science were used to run an industry of outstanding efficiency, whereas the dominant racial ideology of the Nazis was without any scientific basis at all. This of course is perhaps the most extreme example, but similar contradictions exist to some extent in nearly all twentieth-century societies. It is perhaps a partial reflection of this situation that the two most popular philosophical theories of probability of the present time are the ultra-objective propensity theory, which sees probability as part of the material world, and the subjective theory, which makes probability a measure of the personal belief of a particular individual. One of the themes of the present book will be to try to find ways of overcoming this uncomfortably sharp dichotomy.

In the next chapter we shall consider the first of the philosophical views of probability to emerge in the twentieth century – the logical theory. This was developed by Keynes in the Cambridge of the Edwardian era. The logical theory is the one most similar to the traditional classical view, and perhaps the Cambridge of that time gave rise to a late flowering of the ideas and ideals of the age of reason. This flowering was brought to an end by the outbreak of the First World War, which gave such a striking demonstration of that typically twentieth-century combination of irrationality going hand in hand with scientific and technological ingenuity.

3 The logical theory

In the first few decades of the twentieth century the logical theory of probability was developed mainly at Cambridge, though later on it was taken up by members and associates of the Vienna Circle. Carnap supported the logical theory in the 1950s, and Popper too advocated a somewhat different version of the theory at that time. I will mention some of these later developments from time to time in what follows, but, in this chapter, I will focus on Cambridge in the Edwardian era before the First World War, and, in particular, on the work of Keynes. Keynes was not, however, the first, or the only, person to work on the logical theory of probability in Cambridge at that time. He was preceded by W. E. Johnson, whose lectures he attended. These lectures were also attended by Harold Jeffreys, who went on to develop a logical theory of probability which was eventually published in book form in 1939. I have chosen to concentrate on Keynes, however, because, of these authors, he is the one who lays most emphasis on the philosophical aspects which are the subject of the present book. Keynes's work on the philosophy of probability is part of a notable flowering of philosophy which occurred in Cambridge in the Edwardian era, and which involved Bertrand Russell, G. E. Moore and the young Wittgenstein, as well as Keynes. Since Keynes's own work is best understood as part of this philosophical context, I will say something about it in the next section.

Cambridge in the Edwardian era[1]

I will take the period from about 1900 to the outbreak of the First World War as the Edwardian era. This does not quite coincide with the reign of Edward VII, which lasted from 1901 to 1910. However, the fit is good enough for the name to be appropriate. Naturally, the monarch himself did not exert a great deal of influence over the developments we shall be considering, although he did suggest Russell's famous example: 'The King of France is bald'. Indeed Russell writes: 'If we say "the King of England is bald", that is, it would seem, ... a statement ... about the actual man denoted by the meaning. But now consider "the King of France is bald."' (1905:46). This is probably the only allusion to Edward VII in the philosophy of the time, and indeed by the Edwardian era I mean the historical period from about the turn of the century to the outbreak of the First World War – a time which had its own special social, cultural and intellectual characteristics. If a short expression was required to describe this period, the phrase 'the era of paradoxes' could perhaps be used. It was during this time that the paradoxes of logic came to light, and, as we shall see, Keynes laid great stress on some paradoxes of probability. Yet despite the emergence of these paradoxes, the thinkers of this period retained much of the old Enlightenment faith in the power of reason, and they believed that improvements to logic and a better analysis of rationality could overcome the difficulties. This faith was to be rudely shattered by the First World War, the rise of fascism and the great depression of the 1930s.

Keynes published his views on probability in book form in his 1921 *Treatise on Probability.* However, the work for the book had been done in the Edwardian era, and indeed the *Treatise* had been set up in proof in 1913. The outbreak of war delayed its publication, which Keynes was only able to complete after his work for the war, at the peace conference, and on his critique of the Versailles treaty in his 1919 work *The Economic Consequences of the Peace.* Thus, as Skidelsky says: '...the *Treatise* was a pre-war book, reflecting the way pre-war Cambridge did its philosophy.' (1992:56).

Keynes joined King's College Cambridge as an undergraduate in the autumn of 1902, and in February of 1903 he was initiated into a secret society known as the Apostles. The members of this society considered themselves to be, and indeed largely were, the best intellects of Cambridge. It was an elite within an elite. The secrecy was designed to ensure that members could express unorthodox opinions with complete freedom and no fears of social reprisals. Much later the Apostles became discredited when it emerged that several of them had become Russian spies, but when Keynes joined it was at its height and played a crucial role in the intellectual achievements of the time. The only major philosopher at Cambridge who did not get involved with the Apostles was Wittgenstein. He arrived in Cambridge in 1911 to study with Russell, and in November 1912 was duly elected to the Apostles. However, after attending only one meeting, he resigned. Still, Wittgenstein was not a very clubbable man, and, at a later stage of his life, refused to attend even a single meeting of another famous intellectual group – the Vienna Circle. Despite his hostile attitude to the Apostles, Wittgenstein was of course influenced by the intellectual currents of pre-war Cambridge, just as he was later influenced by the Vienna Circle. Indeed Wittgenstein's *Tractatus Logico-Philosophicus* of 1921 contains a sketch of a logical theory of probability (see propositions 4.464 and 5.15–5.156).

When Keynes joined in 1903, the most distinguished Apostles were Bertrand Russell, who had become a member in 1892, and G. E. Moore, who had been initiated in 1894. As these dates indicate, Russell and Moore were about ten years older than Keynes, and they exercised a considerable influence on the development of his thought. Both published a book in 1903. Russell's was *The Principles of Mathematics,* and Moore's was *Principia Ethica.* In his 1938 talk 'My Early Beliefs', Keynes was to say that in his early work on probability: 'I was writing under the joint influence of Moore's *Principia Ethica* and Russell's *Principia Mathematica.*' (1938:445). Let us look first at the influence of Russell.

Between 1903 and 1910 Russell worked on the logicist programme for the foundations of mathematics. This was an attempt to reduce mathematics to logic in the sense of setting up a formal axiomatic deductive system whose axioms would be self-evident truths of logic and within which it would be possible to prove any mathematical theorem. An earlier attempt to carry out this programme had been made by the German logician Frege, but Russell's discovery of a fundamental logical paradox had shown that Frege's system did not work (for more historical details see Gillies 1982:91–3). In his *Principles of Mathematics* of 1903, Russell published this paradox and also his first attempt to overcome the difficulty using what is known as the theory of types. In the next seven years, he developed his logical system, and then, with the help of another Apostle, Whitehead, gave a full account of it in three enormous volumes of formal mathematical logic entitled *Principia Mathematica.* The first volume appeared in 1910, and Russell says: 'Although the third volume of this work was not published until 1913, our part in it (apart from proof-reading) was finished in 1910 when we took the whole manuscript to the Cambridge University Press.' (1959:74).

Russell did not actually have a post at Cambridge University during this period, but he did remain in touch with the Apostles. He was indeed collaborating for much of the time with one of them (Whitehead). It was not until October 1910 that Russell returned to Cambridge as a fellow of Trinity College and lecturer in the principles of mathematics. However, he would undoubtedly have had many discussions with Keynes, even during the period 1903–10. Russell says in his autobiography: 'I first knew Keynes through his father ... Keynes's father taught old-fashioned formal logic in Cambridge ...' (1967:71). Regarding Keynes himself he says: '... I was considerably concerned in his *Treatise on Probability,* many parts of which I discussed with him in detail.' (1967:71). Russell also mentions on the same page that in 1904 Keynes visited him for a weekend in the country. Russell comments later: 'Keynes's intellect was the sharpest and clearest that I have ever known. When I argued with him, I felt that I took my life in my hands, and I seldom emerged without feeling something of a fool.' (1967:72).

It is not difficult to see how Russell's work on logic could have had a strong indirect influence on Keynes. Russell was working out the principles of deductive logic used in mathematics, but what about the reasoning from evidence to hypotheses and predictions characteristic of science and so many everyday considerations? It could be argued that, as well as a deductive logic, one needed an inductive logic to cover such empirical reasoning. Moreover, this inductive logic would be closely connected to, perhaps identical with, probability theory. Part II of Keynes's *Treatise on Probability* is concerned with setting up probability theory as a system of formal logic, and Keynes remarks at the beginning: 'The reader will readily perceive that this Part would never have been written except under the influence of Mr. Russell's *Principia Mathematica.*' (1921:115). Moreover, after finishing *Principia Mathematica,* Russell himself began to take an interest in inductive logic. In his *Problems of Philosophy* of 1912, Chapter VI is about induction, and in it Russell advocates a probabilistic approach to inductive reasoning. We see here a typical case of the influence of members of the same intellectual circle on each other. Keynes writes in the preface to his *Treatise*: 'It may be perceived that I have been much influenced by W. E. Johnson, G. E. Moore, and Bertrand Russell, that is to say by Cambridge, ...' (1921:*v*), while Russell writes in the preface to his *Problems of Philosophy*: 'I have derived valuable assistance from unpublished writings of Mr. G. E. Moore and Mr. J. M. Keynes: ... from the latter as regards probability and induction.'

Although Russell undoubtedly exerted a very considerable influence on Keynes, it seems that the first stimulus which prompted Keynes to work on the foundations of probability came from Moore. According to Skidelsky:

> Keynes's investigation into the meaning of probability was to occupy most of his leisure from 1906 to 1914. But his first discussion on the subject dates from 23 January 1904 when he read a paper to the Apostles entitled 'Ethics in Relation to Conduct'. This makes it clear that his interest arose directly out of the intellectual ferment caused by the appearance of *Principia Ethica*. (1983:152)

This agrees with the reminiscences of Keynes himself who wrote: 'The large part played by considerations of probability in his [i.e. Moore's] theory of right conduct was, indeed, an important contributory cause to my spending all the leisure of many years on the study of that subject' (1938:445). It is somewhat curious that Keynes should describe his *work* on probability as *leisure*! On 12 December 1907, he submitted a dissertation on probability for the prize fellowship competition at King's College Cambridge, but he was unsuccessful. On 17 March 1908, the college awarded fellowships instead to two gentlemen by the names of Dobbs and Page. However, Cambridge was not in the mood to expel its rising star. In June 1908 Keynes was offered a lectureship in economics, and on 16 March 1909 a revised version of his dissertation on probability won him a fellowship at King's. Let us return now to the problem which started these investigations.

In the *Treatise* the problem is discussed in Chapter XXVI (Keynes 1921:307–23), which is entitled 'The Application of Probability to Conduct'. Moore's argument in *Principia Ethica* was along the following lines (see Keynes 1921:309 for an exact quotation). We should act in order to bring about the greatest amount of goodness, but we can only calculate the probable effects of our actions in the 'immediate future'. We really know nothing about their long-term consequences. Moreover, these long-term consequences may be such as to reverse the balance of good produced by our action in the short term. Moore used these sceptical doubts to argue that we can do no better in most cases than to follow the existing rules of morality. Keynes disliked this conclusion, since he believed that a rational member of the Apostles could judge with confidence that some actions contravening conventional morality were nonetheless good. Keynes may have been thinking of homosexual acts, though later members of the Apostles were to judge the action of becoming a Russian spy in this light. Keynes thought that the mistake (as he saw it) in Moore's argument lay in Moore's adopting the wrong interpretation of probability. As he puts it:

> If good is additive, if we have reason to think that of two actions one produces more good than the other in the near future, and if we have no means of discriminating between their results in the distant future, then by what seems a legitimate application of the Principle of Indifference we may suppose that there is a probability in favour of the former action. Mr. Moore's argument must be derived from the empirical or frequency theory of probability, according to which we must know for certain what will happen *generally* (whatever that may mean) before we can assert a probability.
>
> (Keynes 1921:309–10)

We shall consider the Principle of Indifference in more detail later in this chapter, but its application in the present case is a simple one. Let us suppose we are deciding between two courses of action A and B. We can be reasonably sure that in the short term the good produced by A will be greater than the good produced by B. Regarding the long-term consequences of A and B, we have no real knowledge and so are indifferent between the two possibilities: (a) the good produced by A long term will be greater than the good produced by B long term, and (b) the good produced by B long term will be greater than the good produced by A long term. Given this indifference we should assign possibilities (a) and (b) equal probabilities. The desire to maximise expected good now leads us to prefer action A. The general conclusion is that we should carry out the action which produces the most goodness in the short term, even if this contradicts the rules of conventional morality. It is interesting to note the similarity of this to Pascal's wager, discussed above (p. 12) – even though Keynes reaches a conclusion which is the direct opposite of Pascal's.

Another observation worth making is that these ethical arguments of the young Keynes have quite a close connection with his later discussions of investment. We have only to substitute for a moral individual wondering what action will produce the greatest amount of good a business man wondering what investment will bring him the greatest amount of profit. Once again the business man can only reasonably calculate the short-term profits of his investment, and these might in some cases be outweighed by long-term losses.

This concludes my account of the intellectual milieu in which Keynes's ideas on probability took shape. In the next section I will begin my exposition of the ideas themselves. In describing these ideas in detail, it will be possible to give some further examples of the influence on Keynes of his mentors – Moore and Russell.

Probability as a logical relation

In the case of deductive logic a conclusion is entailed by the premises, and it is certain given those premises. Thus, if our premises are that all ravens are black and George is a raven, it follows with certainty that George is black. But now let us consider an inductive, rather than deductive, case. Suppose our premises are the evidence (e say) that several thousand ravens have been observed, and that they were all black. Suppose further that we are considering the hypothesis (h say) that all ravens are black, or the prediction (d say) that the next observed raven will be black. Hume argued, and this is in agreement with modern logic, that neither h nor d follow logically from e. Yet even though e does not entail either h or d, could we not say that e *partially entails* h and d, since e surely gives some support for these conclusions? This line of thought suggests that there might be a logical theory of partial entailment which generalises the ordinary theory of full entailment which is found in deductive logic. This is the starting point of Keynes's approach to probability. He writes:

> Inasmuch as it is always assumed that we can sometimes judge directly that a conclusion *follows from* a premiss, it is no great extension of this assumption to suppose that we can sometimes recognise that a conclusion *partially follows from,* or stands in a relation of probability to a premiss.
>
> (Keynes 1921:52)

and again:

> We are claiming, in fact, to cognise correctly a logical connection between one set of propositions which we call our evidence and which we suppose ourselves to know, and another set which we call our conclusions, and to which we attach more or less weight according to the grounds supplied by the first.... It is not straining the use of words to speak of this as the relation of probability.
>
> (1921:5–6)

So a probability is the degree of a partial entailment.

One immediate consequence of this approach is that it makes all probabilities conditional. We cannot speak simply of the probability of a hypothesis, but only of its probability relative to some evidence which partially entails it. Keynes puts the point as follows:

> No proposition is in itself either probable or improbable, just as no place can be intrinsically distant; and the probability of the same statement varies with the evidence presented, which is, as it were, its origin of reference.
>
> (1921:7)

At first this would seem to conflict with our ordinary use of the probability concept, for we do often speak simply of the probability of some outcome. Keynes would reply that in such cases a standard body of evidence is assumed.

So far the probability relation has been described as 'degree of partial entailment', but Keynes gives another account of it in the following passage:

> *Let our premisses consist of any set of propositions h*, *and our conclusion consists of any set of propositions a*, *then, if a knowledge of h justifies a*

> *rational belief in a of degree α, we say that there is a* probability-relation *of degree α between* a *and h.'*
>
> (1921:4)

Here Keynes makes the assumption that if *h* partially entails *a* to degree *α*, then given *h* it is rational to believe *a* to degree α. To put it less formally, he identifies 'degrees of partial entailment' and 'degrees of rational belief.' This assumption seems at first sight plausible, but it has been challenged by Popper. One of Popper's arguments is the following. Suppose we have finite evidence and a generalisation which may have a potentially infinite number of instances, for example the e and h in the ravens example given earlier (pp. 29–30). Now here h goes so to speak infinitely beyond e, and thus, Popper argues, the degree to which e partially entails h is zero. This conclusion was also accepted by Carnap. But, Popper's argument continues, although the degree to which finite evidence partially entails a universal generalisation is zero, it may nonetheless be possible to have a non-zero degree of rational belief in a universal generalisation given finite evidence. Indeed, this is often the case when we entertain some finite degree of rational belief in a scientific theory. So, Popper concludes, we should not identify degree of partial entailment with degree of rational belief. Popper accepts a logical interpretation of probability where probability is identified with degree of partial entailment, but, since these degrees of partial entailment are no longer degrees of rational belief, his theory differs from that of Keynes. Popper identifies degree of rational belief with what he calls 'degree of corroboration', and so sums up his position as follows:

> *we may learn from experience more and more about universal laws without ever increasing their probability*; ... we may test and corroborate some of them better and better, thereby increasing their *degree of corroboration* without altering their *probability* whose value remains zero.
>
> (Popper 1959a:383)

I give a more detailed discussion of this argument in Gillies (1988a:192–5) but will now continue my exposition of Keynes.

The next question which might be asked regarding Keynes's approach is the following: 'how do we obtain knowledge about this logical relation of probability, and, in particular, how are the axioms of probability theory to be established from this point of view?' On the general problem of knowledge Keynes adopted a Russellian position. Russell held that some of our knowledge is obtained directly or 'by acquaintance'. His views on what we could know in this way varied, but the set always included our immediate sense perceptions. The rest of our knowledge is 'knowledge by description' and is ultimately based on our 'knowledge by acquaintance'. In analysing the relations between the two sorts of knowledge, Russell thought that his theory of descriptions could play an important rôle. In Russellian vein, Keynes writes: 'About our own existence, our own sense-data, some logical ideas, and some logical relations, it is usually agreed that we have direct knowledge.' (1921:14). Though he adds later: 'Some men –

indeed it is obviously the case – may have a greater power of logical intuition than others.' (1921:18). In particular, we get to know at least some probability relations by direct acquaintance or immediate logical intuition. As Keynes says: 'We pass from a knowledge of the proposition *a* to a knowledge about the proposition *b* by perceiving a logical relation between them. With this logical relation we have direct acquaintance.' (1921:13). Indeed, Keynes appears to argue at times that all logical relations are known by direct acquaintance. Thus he says: 'When we know something by argument this must be through direct acquaintance with some logical relation between the conclusion and the premiss.' (1921:14).

This view is, however, rather extreme, since it seems to make the axiomatisation of logic or probability unnecessary. Keynes does indeed recognise this at other points. Thus he writes:

> While we may possess a faculty of direct recognition of many relations of probability, as in the case of many other logical relations, yet some may be much more easily recognisable than others. The object of a logical system of probability is to enable us to know the relations, which cannot be easily perceived, by means of other relations which we can recognise more distinctly – to convert, in fact, vague knowledge into more distinct knowledge.
>
> (Keynes 1921:53)

This approach underlies Keynes's attempt in Part II of the *Treatise* to present a formal axiomatic system for probability. He says that the object of this part of his book is:

> ... to show that all the usually assumed conclusions in the fundamental logic of inference and probability follow rigorously from a few axioms, in accordance with the fundamental conceptions expounded in Part I. This body of axioms and theorems corresponds, I think, to what logicians have termed the *Laws of Thought,* when they have meant by this something narrower than the whole system of formal truth. But it goes beyond what has been usual, in dealing at the same time with the laws of probable, as well as of necessary, inference.
>
> (Keynes 1921:133)

As already remarked, Keynes's approach to probability here is exactly the same as that of Russell and Whitehead to deductive logic. The aim of *Principia Mathematica* was to start from axioms which were obviously correct to logical intuition, and from these to deduce results which were thereby shown to be logically valid but which might not be so immediately obvious to logical intuition. Unfortunately, some of the axioms which Russell and Whitehead used in *Principia Mathematica* were far from being obviously correct to the intuitions of many mathematicians, and, as we shall in the next chapter, similar criticisms were directed against Keynes's *Treatise.*

For Keynes probability was degree of *rational* belief *not* simply degree of belief. As he says:

> ... in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. The Theory of Probability is logical, therefore, because it is concerned with the degree of belief which it is *rational* to entertain in given conditions, and not merely with the actual beliefs of particular individuals, which may or may not be rational.
>
> (Keynes 1921:4)

Here Keynes speaks of probabilities as being fixed objectively, but he is not using objective here in the way we have defined it to refer to things in the material world. He means objective in the Platonic sense, referring to something in a supposed Platonic world of abstract ideas. Indeed, Keynes goes so far as to suggest that probability relations which none of us will ever be able to apprehend may nonetheless exist in the Platonic world. He writes: 'The perceptions of some relations of probability may be outside the powers of some or all of us.' (Keynes 1921:18).

In his later reminiscences, Keynes does refer to the faith of his group at that time as: 'some sort of relation of neo-Platonism' (1938:438). We can see here clearly the influence of G. E. Moore. In his *Principia Ethica* Moore had argued that good was a non-natural property which could be known only by intuition. In the same way Keynes argues that probabilities are logical relations which are known by intuition.[2] In fact, there is a very notable similarity between the Platonic world as postulated by Cambridge philosophers in the Edwardian era and the Platonic world as originally described by Plato. Plato's world of objective ideas contained the ethical qualities with the idea of the Good holding the principal place, but it also contained mathematical objects. The Cambridge philosophers thought that they had reduced mathematics to logic. So their Platonic world contained, as well as ethical qualities such as 'good', logical relations. These similarities perhaps reflect a similarity in the social basis of the thought. Plato and his circle were an elite group of wealthy intellectuals who discussed philosophy in the grove of the hero Academus, not far from the great commercial city of Athens. The Apostles were an elite group of wealthy intellectuals who discussed philosophy in the pleasant surroundings of Cambridge, not far from the great commercial city of London.

Measurable and non-measurable probabilities: the Principle of Indifference

In the usual mathematical treatments of probability, all probabilities are regarded as having a definite numerical value in the interval [0, 1]. Keynes, however, does not think that all probabilities have a numerical value. On the contrary, some probabilities may not even be comparable. As he says:

> ... no exercise of the practical judgment is possible, by which a numerical value can actually be given to the probability of every argument. So far from our being able to measure them, it is not even clear that we are always able to place them in an order of magnitude. Nor has any theoretical rule for their evaluation ever been suggested.
>
> (Keynes 1921:27–8)

So if we have two probabilities, a variety of situations can hold. They may both have a numerical value. Again, although we may not be able to assign a numerical value to both of them, we might perhaps be able to say that one is greater than the other. In still other cases we may not be able to make any comparison. As Keynes puts it:

> I maintain, then, in what follows, that there are some pairs of probabilities between the members of which *no* comparison of magnitude is possible; that we can say, nevertheless, of some pairs of relations of probability that the one is greater and the other less, although it is not possible to measure the difference between them; and that in a very special type of case, to be dealt with later, a meaning can be given to a *numerical* comparison of magnitude.
>
> (Keynes 1921:34)

The set of probabilities is thus not linearly ordered. It has, however, a special kind of partial ordering which is illustrated in Figure 3.1.

Keynes comments on the diagram in Figure 3.1 as follows:

> *O* represents impossibility, *I* certainty, and *A* a numerically measurable probability intermediate between *O* and *I*; *U, V, W, X, Y, Z* are non-numerical probabilities, of which, however, *V* is less than the numerical probability *A,* and is also less than *W, X,* and *Y. X* and *Y* are both greater than *W,* and greater than *V,* but are not comparable with one another, or with *A*. *V* and *Z* are both less than *W, X,* and *Y,* but are not comparable with one another; *U* is not quantitatively comparable with any of the probabilities *V, W, X, Y, Z.*
>
> (Keynes 1921:39)

Attitudes towards this rather complicated construction differ considerably. De Finetti regarded Keynes's non-numerical probabilities as an unfortunate departure from the simplicity and power of the mathematical theory of probability. He wrote:

> ... for Keynes there also exist ... probabilities which cannot be expressed as numbers.
>
> Keynes's position is certainly not suited to the development of a mathematical probability theory and is also hardly in keeping with the intuitive idea of probability....

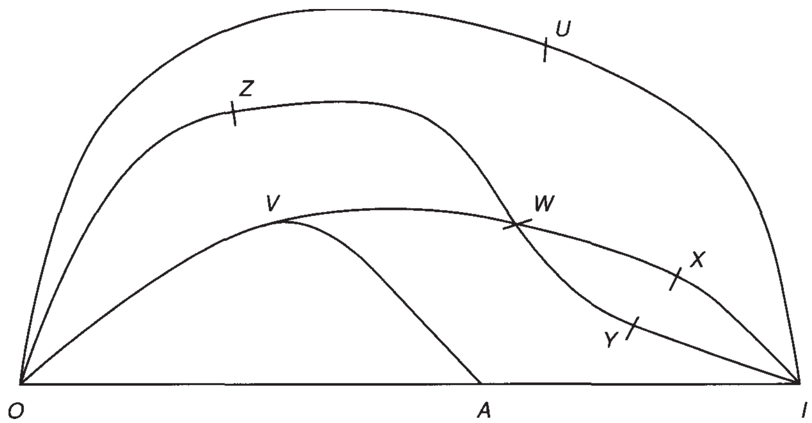


Figure 3.1 Partial order of the set of probabilities in Keynes's logical theory (Keynes 1921:39)

> ... I myself regard as unacceptable, as a matter of principle, Keynes's position (the more so since the reservations which he had disappear when one adopts a subjective point of view).
>
> (De Finetti 1938:359)

Runde (1994), on the other hand, sees Keynes's qualitative approach to probability as more realistic than a numerical approach in many cases. Indeed, Runde argues that we can preserve Keynes's non-numerical theory of probability while abandoning Keynes's Platonism and reliance on intuition.

What then are the cases in which numerical values can be assigned to probabilities? Keynes answers unequivocally: 'In order that numerical measurement may be possible, we must be given a number of *equally* probable alternatives.' (1921:41). He even claims that this is something on which all probabilists agree: 'It has always been agreed that a numerical measure can actually be obtained in those cases only in which a reduction to a set of exclusive and exhaustive *equiprobable* alternatives is practicable.' (1921:65).

So in order to get numerical probabilities we have to be able to judge that a number of cases are equally probable, and to enable us to make this judgement we need an a priori principle. This a priori principle is called by Keynes the Principle of Indifference. The name is original to him but the principle itself, he says, was introduced by J. Bernoulli under the name of the Principle of Non-sufficient Reason. Keynes gives the following preliminary statement of the principle:

> The Principle of Indifference asserts that if there is no *known* reason for predicating of our subject one rather than another of several alternatives, then relatively to such knowledge the assertions of each of these alternatives have an equal probability.
>
> (Keynes 1921:42)

Unfortunately the Principle of Indifference leads to a number of paradoxes, some of which we shall consider in the next section. Before examining these objections to the principle, however, it will be interesting to see how it is applied in the Bayesian approach. This will constitute a brief introduction to Bayesianism, a theory to which we will return from time to time in what follows. Our sketch below illustrates Bayesianism of the logical variety. Nowadays, subjective Bayesianism is much more popular. However, the difference between logical and subjective Bayesianism will become clear when we consider the subjective theory of probability in the next chapter.

Let us return then to our simple example of the ravens. Let h be the hypothesis that all ravens are black. Let e_1 = the first observed raven was black, ..., e_n = the nth observed raven was black, and e (our evidence) = e_1 & e_2 & ... & e_n. h does not follow logically from e, but, if we accept the logical interpretation of probability, it makes sense to say that e makes h probable to some degree. The conditional probability of h given e is written P(h | e). The aim of the Bayesian school is to find methods for calculating P(h | e). This, the Bayesians think, would provide an underpining for the inductive inferences from evidence to hypothesis which occur in both science and everyday life.

According to Axiom 3, which I will present in the next chapter,

$$P(h \mid e) = \frac{P(e \ \& \ h)}{P(e)} \quad \text{provided } P(e) \neq 0 \tag{3.1}$$

It follows that, under the same condition,

$$P(h \mid e) = \frac{P(e \mid h)P(h)}{P(e)} \tag{3.2}$$

This is a simple version of Bayes's theorem. It is worth explaining its various components.

P(h | e) is known as the posterior probability of h given e.
P(e | h) is known as the likelihood.
P(h) is the prior probability of h.
P(e) is the prior probability of e.

The idea of Bayesian inference is to use Bayes's theorem to go from the prior probability of a hypothesis h to its posterior probability in the light of evidence. The change from P(h) to P(h | e) is known as *Bayesian conditionalisation.* In the logical version of Bayesianism, this procedure involves the use of the Principle of Indifference as I will now explain.

In order to calculate P(h | e), we have to evaluate all the elements on the right hand side of Equation 3.2. Likelihoods are often easy to calculate. In fact, if e follows logically from h, then P(e | h) = 1. If h is a statistical hypothesis, an easy probability calculation often gives P(e | h). Thus we can calculate P(h | e), if we can evaluate P(h) and P(e); but how can this be done?

One method is the following. Suppose $h = h_1$ *where we have* m *possible mutually exclusive hypotheses* $h_1, h_2, \ldots, h_m$ *(say in this case specifying different colours for the ravens), and suppose further that we have no a priori reason for preferring* h_i *to* h_j *where* $i \neq j$. *Then by the Principle of Indifference*

$$P(h_1) = P(h_2) = \ldots = P(h_m) = 1/m$$

Further from a theorem of the probability calculus

$$P(e) = \sum_{i=1}^{m} P(e \mid h_i) P(h_i)$$

Therefore substituting we get

$$P(h \mid e) = \frac{P(e \mid h)}{\sum_{i=1}^{m} P(e \mid h_i)}$$

So, if we can evaluate the likelihoods, which, as we remarked, is often easy, we can compute P(h | e). In this way the logical Bayesians hope to calculate the degrees of rational belief in hypotheses given evidence, or, as it is sometimes put, to construct an inductive logic. This programme is an attractive one, but it does depend on using the Principle of Indifference, which, as we shall see in the next section, is fraught with difficulties.

Paradoxes of the Principle of Indifference

The trouble with the Principle of Indifference is that it gives rise to a number of paradoxes. These contradictions were discovered over quite a long period of time. The earliest seems to have been the needle problem, which Buffon published in 1733.[3] Further paradoxes were published by a number of authors including, notably, Bertrand (1889) and Borel (1909).[3] It is greatly to Keynes's credit that, although he advocates the Principle of Indifference, he gives the best statement in the literature (Keynes 1921: Chapter 4) of the paradoxes to which it gives rise. In this section I will present three of the paradoxes and give a generalisation of the last two. Then in the next section, I will examine attempts to solve the paradoxes, introducing in the course of discussion yet another paradox.

The first of the paradoxes is called the book paradox. Consider a book in a specified place in a library. Let us suppose that we have never visited the library or seen a copy of the book. So we have no idea what the colour of its cover is. In these circumstances it could be argued that we have no more reason to suppose that the cover is red than that it is not red. Thus, using the Principle of Indifference, we have P(red) = $^1/_2$. Similarly, however, P(blue), P(green) and P(yellow) are all $^1/_2$, which contradicts the principle of the probability calculus that the sum of mutually exclusive possibilities must be less than or equal to 1. The book paradox is perhaps not so difficult to resolve, but its discussion leads to some interesting points and it is included for that reason. Let us now turn to a more problematic case. This is the wine–water paradox.

Suppose we have a mixture of wine and water and we know that at most there is 3 times as much of one as of the other, but nothing more about the mixture. We have

$$^1/_3 \leq \text{wine/water} \leq 3$$

and by the Principle of Indifference, the ratio of wine to water has a uniform probability density in the interval [$^1/_3$, 3]. Therefore

$$\begin{aligned} P(\text{wine / water} \leq 2) &= \frac{2 - 1/3}{3 - 1/3} \\ &= 5/8 \end{aligned}$$

But also

$$^1/_3 \leq \text{water/wine} \leq 3$$

and by the Principle of Indifference, the ratio of water to wine has a uniform probability density in the interval [$^1/_3$, 3]. Therefore

$$\begin{aligned} P(\text{water / wine} \geq 1/2) &= \frac{3 - 1/2}{3 - 1/3} \\ &= 15/16 \end{aligned}$$

But the events 'wine/water ≤ 2' and 'water/wine ≥ $^1/_2$' are the same, and the Principle of Indifference has given them different probabilities.

Our third example of a paradox of the Principle of Indifference belongs to a class known as the paradoxes of geometrical probability, because they involve calculating the probabilities of various geometrical figures. The present example was published by Bertrand in 1889. Consider a fixed circle and select a chord at random. What is the probability that this random chord is longer than the side of the equilateral triangle inscribed in the circle? We can use the Principle of Indifference in three plausible ways to produce three different values for this probability, which can be conveniently abbreviated to P(CLSE) (= the probability that the chord is longer than the side of the equilateral triangle inscribed in the circle). Let us begin by considering (Figure 3.2) an equilateral triangle XYZ inscribed in a circle with centre O whose radius we will suppose to be *R*.

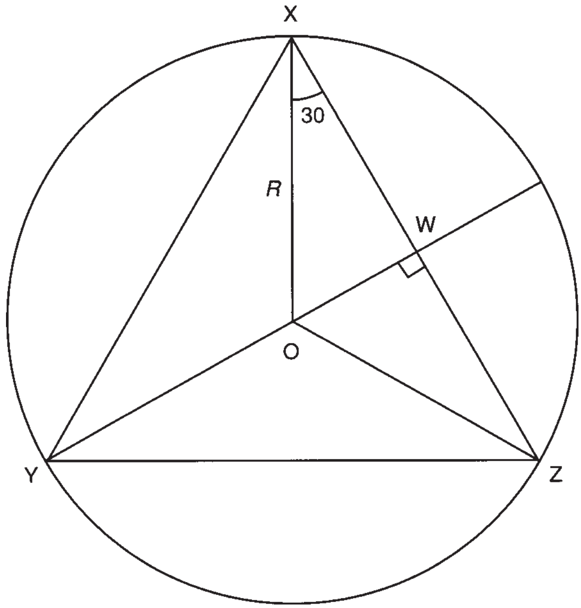


Figure 3.2 An equilateral triangle inscribed inside a circle with centre O and radius *R*

Extend YO to meet XZ at W. Then OWZ is a right angle, and XW = WZ. Moreover OW = $R\sin 30 = R/2$. We can now use these geometrical facts for our first calculation (Figure 3.3).

Let AB be our random chord and OW the perpendicular from O meeting the circle again at C. AB is longer than the side of the inscribed equilateral triangle if OW < $R/2$. But we have no reason to suppose that W is at any point on OC rather than any other point. So, by the Principle of Indifference, OW has a uniform probability density in the interval $[0, R]$. Therefore

$$\text{P(CLSE)} = \text{P}(\text{OW} < R/2) = {}^1\!/_2$$

For our next calculation, let us turn to Figure 3.4.

Once again AB is the random chord. Let AA′ A″ be the inscribed equilateral triangle with A as one of its vertices. Draw the tangent at A, and let θ be the angle between this tangent and AB. Then AB is longer than the side of the inscribed equilateral triangle if $60 < \theta < 120$. We have no reason to suppose that θ has any value between 0 and 180 rather than any other. So, by the Principle of Indifference, θ has a uniform probability distribution in the interval [0, 180]. Therefore

$$\text{P(CLSE)} = \text{P}(60 < \theta < 120) = 1/3$$

For our third calculation, consider Figure 3.5.

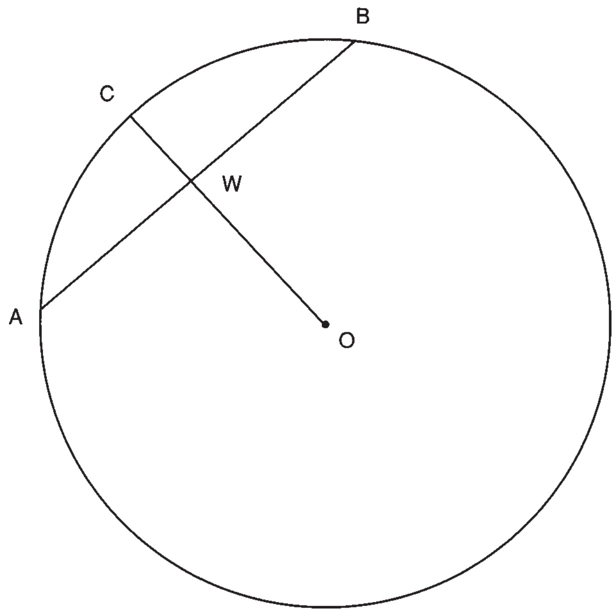


Figure 3.3 First calculation of P(CLSE)

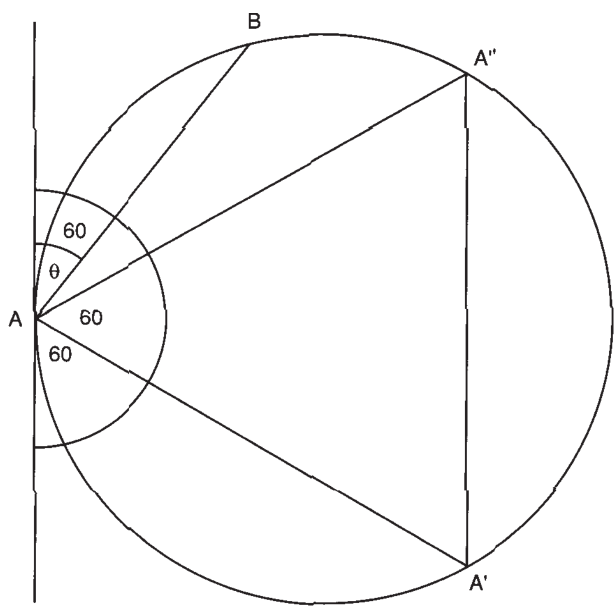


Figure 3.4 Second calculation of P(CLSE)

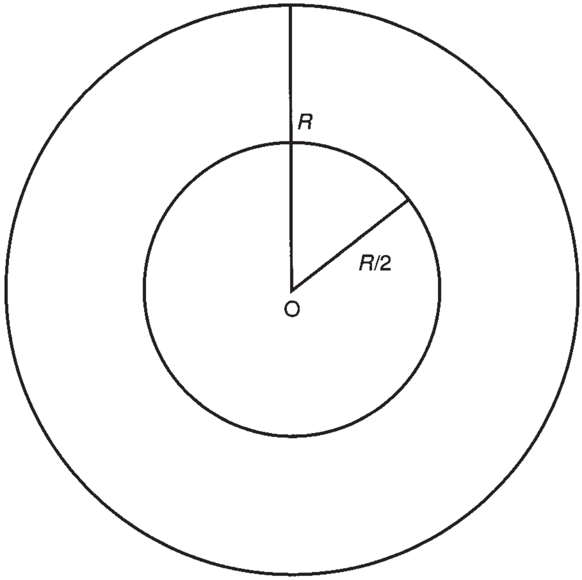


Figure 3.5 Third calculation of P(CLSE)

Here we inscribe inside our principal circle a small circle with the same centre O, but half the radius ($R/2$). The random chord AB will be longer than the side of the inscribed equilateral triangle if its centre W lies inside this small circle. We have no reason to suppose that W lies at any point in the main circle rather than any other. So, by the Principle of Indifference, W has a uniform probability density in the main circle. Therefore

$$\mathrm{P(CLSE)} = \frac{\text{Area of small circle}}{\text{Area of main circle}} = \frac{\pi R^2 / 4}{\pi R^2} = 1/4$$

This is a neat example since the value of CLSE is given by successive applications of the Principle of Indifference as $1/2$, $1/3$ and $1/4$.

It is easy to see how we can generalise from the last two examples to produce a paradox in any case which concerns a continuous parameter (θ say) which takes values in an interval $[a, b]$. All we have to do is to consider $\phi = f(\theta)$, where f is a continuous and suitably regular function defined in the interval $[a, b]$ so that $a \leq \theta \leq b$ is logically equivalent to $f(a) \leq \phi \leq f(b)$. If we have no reason to suppose that ? is at one point of the interval $[a, b]$ rather than another, we can then use the Principle of Indifference to give θ a uniform probability density in $[a, b]$. However, we have correspondingly no reason to suppose that f is at one point of the interval $[f(a), f(b)]$ rather than another. So it seems we can equally well use the Principle of Indifference to give ϕ a uniform probability density in $[f(a), f(b)]$. However, the probabilities based on θ having a uniform probability density will in general be different from those based on ϕ having a uniform probability density; and thus the

Principle of Indifference leads to contradictions. The wine–water paradox is a simple example of this, for if we set θ = wine/water then the paradox is generated by considering $\phi = f(\theta) = 1/\theta$. This concludes my account of the paradoxes of the Principle of Indifference. Let us next examine the attempts that have been made to solve them.

Possible solutions to the paradoxes

Let us start with the book paradox. Here it was argued that we have no more reason to suppose that the book is red than that it is not red, so that by the Principle of Indifference P(red) = $^1/_2$. However, the premise used is highly doubtful. The alternative not-red can be divided into blue and not-(red or blue), and blue is of the same form as red. Thus the alternatives red and not-red are not suitable for the application of the Principle of Indifference. Indeed, it seems obvious that the alternative not-red is more probable than the alternative red. A similar example where the Principle of Indifference does seem to be applicable is the following.[4] Suppose we are considering the colour of a car concerning which we know only its year and type. From a catalogue we learn that cars of that year and type were produced in one of seven different colours. In this case it does seem reasonable to use the Principle of Indifference to assign each of these colours the probability $^1/_7$.

We can now generalise from this example to explain the way in which Keynes seeks to solve the paradoxes. His idea is that we should only apply the Principle of Indifference to cases where the alternatives are finite in number and 'indivisible'. As he says:

> Let the alternatives, the equiprobability of which we seek to establish by means of the Principle of Indifference, be $\phi(a_1)$, $\phi(a_2)$, ..., $\phi(a_r)$, and let the evidence be h. Then it is a necessary condition for the application of the principle, that these should be, relatively to the evidence, *indivisible* alternatives of the form $\phi(x)$.
>
> (Keynes 1921:60)

In the book paradox, we cannot apply the Principle of Indifference because one of the alternatives, i.e. not-red, is, as we have seen, divisible, into sub-alternatives of the same form as the other alternative (red). In the car example, however, the alternatives are all indivisible of the form: possible colour of a car of that year and type. So the Principle of Indifference can legitimately be applied.

The trouble with this suggestion is that it appears to rule out applying the Principle of Indifference to any continuous case in which a parameter θ lies somewhere in an interval $[a, b]$. In such cases either θ is considered as having an infinite number of values, or, if we divide the interval into a finite number of sub-intervals, these sub-intervals can always be divided into further sub-intervals. It looks as if Keynes's modification of the Principle of Indifference prevents it being applied to either of these alternatives; but Keynes himself argues that the second alternative can lead to legitimate applications of the principle. He writes:

> Suppose, for instance, that a point lies on a line of length *m.l,* we may write the alternative 'the interval of length l on which the point lies is the xth interval of that length as we move along the line from left to right' $\equiv \phi(x)$; and the Principle of Indifference can then be applied safely to the m alternatives $\phi(1), \phi(2) \ldots \phi(m)$, the number m increasing as the length l of the intervals is diminished. There is no reason why l should not be of any definite length however small.
>
> (Keynes 1921:62)

Keynes's procedure here seems distinctly doubtful. First of all he builds a definite value of the length l into the form of the alternatives. Surely, however, sub-intervals of different lengths have essentially the same form. Moreover, having done this, he allows the length of the sub-intervals to diminish and become as small as we like. More seriously, this approach does not appear to avoid the wine–water paradox. Suppose in that example we divide the interval $[{}^1/_3, 3]$ into n equal sub-intervals $I_1, \ldots, I_n$ and consider the event E that there is less than twice as much wine as water. By taking the length of I_i sufficiently small and representing E first as a combination of events of the form wine/water ε I_j and then of events of the form water/wine ε I_k, we obtain by suitably modifying the previous argument two different probabilities for E.

I conclude that Keynes's modification of the Principle of Indifference renders it inapplicable to the continuous case. This is a severe, perhaps in itself fatal, blow to his logical theory of probability, since many of the most important applications of the mathematical theory of probability involve numerical probabilities with continuous parameters. A philosophical account of probability which excludes such cases can hardly be regarded as adequate. Moreover, it is not at all clear that Keynes's modification even deals adequately with some paradoxes which arise in the simple case of a finite number of discrete alternatives. To see this I will present one such paradox which has played an important rôle in the history of probability theory.

In fact, Bayes in the paper in which he introduces Bayesianism makes an implicit use of the Principle of Indifference at one point. Bayes had doubts about his famous paper, and so it was only published after his death by his friend Price. As Price added an important introduction and appendix to the paper, it seems to me fair to regard it as a joint paper, and I will refer to it as Bayes and Price 1763. Price mentions Bayes's own doubts about his paper in the following passage from his introduction:

> But he [Bayes] afterwards considered, that the *postulate* on which he had argued might not perhaps be looked upon by all as reasonable; and therefore he chose to lay down in another form the proposition in which he thought the solution of the problem is contained, and in a *scholium* to subjoin the reasons why he thought so, rather than to take into his mathematical

> reasoning anything that might admit dispute. This, you will observe, is the method he has pursued in this essay.
>
> (Bayes and Price 1763:134)

Bayes does indeed begin by considering a specific example: his billiard table example. His mathematical analysis of this example is such as would be accepted by Bayesian and non-Bayesian alike.[5] The trouble comes when he generalises from this specific example to the general case of an event M which may or may not occur during a particular trial and about which we know nothing else. Bayes argues in his scholium that the same rule which he derived in the billiard table example applies to such an event, and in doing so he makes an implicit application of the Principle of Indifference.

> And that the same rule is the proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. For, on this account, I may justly reason concerning it as if its probability had been at first unfixed, and then determined in such a manner as to give me no reason to think that, in a certain number of trials, it should happen any one possible number of times than another. But this is exactly the case of the event M.
>
> (Bayes and Price 1763:143)

Bayes is considering an event M about which we know only that it may or may not occur on each of a number (n say) of trials. He argues that there is no reason to suppose that in these trials the event will occur one possible number of times, r say, rather than another, s say, where $r \neq s$ and $0 \leq r, s \leq n$. He then implicitly uses the Principle of Indifference to assign equal probabilities to each of these possible outcomes, and so obtains

$$\text{P(M occurs exactly } r \text{ times in } n \text{ trials)} = 1/(n + 1)$$

We see then that Bayes applies the Principle of Indifference to the number of successes (occurrences of the event M). However, Edwards points out in his commentary on the arguments of Thomas Bayes that we could equally well apply the Principle of Indifference to the possible sequences of successes and failures, and would in that case obtain a different result (see Edwards 1978:118).

To compare the results of the two different applications of the Principle of Indifference, let us write 1 for a success (an occurrence of M), and 0 for a failure, and let us consider the case where $n = 2$. We have four possible sequences of successes and failures, namely 00, 01, 10, 11. Applying Edwards's method, these are each assigned the probability $^1/_4$. If we denote this probability distribution by P, we have P(01 or 10) = $^1/_2$. Applying Bayes's method, there are three possible numbers of successes, namely 0, 1 or 2, and each of these is assigned the probability $^1/_3$. If we denote this probability distribution by P^*, we have P^*(01 or 10) = $^1/_3$. Here then we have a typical paradox of the Principle of Indifference, but here we are dealing not with the continuous case but with the simpler case of a finite number of discrete alternatives. Moreover, the example is not an arbitrary one, but it is very important for the analysis of inductive reasoning. Let us therefore consider whether Keynes's suggested modification of the Principle of Indifference can deal with the problem.

Keynes's notion of indivisible alternatives of the same form is not entirely precise, and there seem to be two ways in which it could be applied in this example. First of all we could say that the possible sequences and the possible numbers of successes are both indivisible alternatives, though of different forms. This would allow both P and P^* as valid. Alternatively, we could argue that the alternatives considered by Bayes are not really indivisible. For example, the alternative of one success in two trials is really divisible into two subalternatives, namely 01 and 10. If we adopted this approach, then P would be valid but P^* invalid. Neither of these two applications of Keynes's approach is, however, satisfactory. The first one allows both P and P^*, and so does not solve the paradox at all. The second eliminates P^*, but there were very good reasons why Bayes wanted to adopt P^* and eliminate P. I will now explain these reasons, and they will show that the second way of applying Keynes's method unfortunately eliminates what, from the point of view of inductive logic, is the wrong alternative.

The key point is that if we adopt P, then learning from experience by Bayesian conditionalisation becomes impossible. To see this let us consider again Equation 3.1, namely

$$P(h \mid e) = \frac{P(e \ \& \ h)}{P(e)} \quad \text{provided } P(e) \neq 0$$

Here let us suppose that e gives the result of the first n trials, and that h is the hypothesis that M occurs on the n + 1th trial. P(h) = $^1/_2$. The posterior probability of h can now be calculated from the above formula. In n trials there are 2^n possible sequences of successes and failures, e is a particular such sequence, and so P(e) = 2^{-n}. Similarly, P(e & h) = $2^{-(n+1)}$. So

$$P(h \mid e) = \frac{2^{-(n+1)}}{2^{-n}} = \tfrac{1}{2}$$

Thus, the posterior probability is the same as the prior probability, and, in the Bayesian framework, no learning from experience can occur.

We see then that Bayes was wise to choose P^* rather than P, and similar choices have been made by later Bayesians. For example, Carnap (1950) considers two confirmation functions $c^\dagger$ and c^*. $c^\dagger$ is obtained by giving equal probabilities to his state descriptions, and c^* by giving equal probabilities to his structure descriptions. Thus, these functions are analogues within Carnap's system to our P and P^*. Carnap (1950:562–5) makes the point that if we adopted $c^\dagger$ then learning from experience would become impossible, and so argues for c^*. Although this is quite reasonable, it is at the same time *ad hoc.* From the point of view of the Principle of Indifference, even with Keynes's modification, there does not seem to be any reason for preferring P^* to P. On the contrary, Keynes's approach suggests that, if anything, we should prefer P to P^*. I can only conclude that Keynes does not succeed in giving a satisfactory solution to the paradoxes of the Principle of Indifference.

Let us next examine some approaches to the paradoxes which are different from Keynes's. We have seen that a group of paradoxes can arise in the continuous case by transforming a parameter θ to another $\phi = f(\theta)$, and then applying the Principle of Indifference first to θ and then to ϕ, thereby producing different results. In some cases an attempt could be made to block this transformation procedure by arguing that one particular parameter was natural for that problem, and that to transform it into another would be bizarre and artificial. Thus, for example, if the parameter were height (h), it might be logically possible to use instead $g = 1/h$, but such a procedure would be peculiar to say the least. The values of h are given directly by the measuring procedure, and within a broad range one has a constant error (ε say), so that the result is $h \pm \varepsilon$. If one used $g = 1/h$ instead of h, it would be necessary to transform the result obtained directly from the measuring procedure, and including the error it would become approximately $g \pm \varepsilon g^2$. In other words, the magnitude of the error would vary with the size of g, which is hardly very convenient. It could thus be argued that considering transformations such as $1/h$ is simply inappropriate, and that therefore in a problem to do with heights the Principle of Indifference could be applied legitimately to h, but not to a transformation of h. I am sympathetic to this approach, which could certainly deal with a number of the paradoxes, but, like so many of the suggested solutions, it does not deal with all the paradoxes. For example, in the wine–water paradox the two parameters considered, namely wine/water and water/wine, are quite symmetrical and equally natural. So this paradox cannot be resolved along the lines just suggested.

Jaynes (1973) wrote a very interesting paper on our third (geometrical) paradox about the random chord of a circle. Most unusually, he argued that one of the three solutions was correct and the other two were wrong. The solution he defended was the first one, i.e. P(CLSE) = $^1/_2$. Jaynes argued that the solution to the problem should satisfy some invariance principles. In particular, if we require that the solution be rotation invariant, scale invariant and translation invariant, this eliminates the second two solutions and leaves the first solution as the only possible one. Using the principles of this first solution he calculated the entire probability distribution of the chord lengths. He and Dr Charles E. Tyler then performed an experiment which consisted of tossing broom straws from a standing position onto a 5-inch-diameter circle drawn on the floor. The results of 129 successful tosses confirmed his calculated distribution: 'with an embarrassingly low value of chi-squared' (Jaynes 1973:487). My reaction to this ingenious proposal is the same as to the previous suggestion. There is no doubt that an appeal to invariance principles can solve some of the paradoxes of the Principle of Indifference in a plausible fashion. Jaynes's paper definitely shows this to be the case. However, invariance principles cannot solve all the paradoxes. In particular they cannot deal with the wine–water paradox, since, as regards invariance, there is nothing to choose between the parameters wine/water and water/wine, as Jaynes (1973:490) himself says.

Jaynes has, however, a very interesting general argument in favour of the Principle of Indifference. His point is that this principle has often been used with great success in physics, and so cannot be altogether valueless. He illustrates this with the example of the viscosity of a gas:

> For example, given the average particle density and total energy of a gas, predict its viscosity. The answer, evidently, depends on the exact spatial and velocity distributions of the molecules (in fact, it depends critically on position-velocity correlations), and nothing in the given data seems to tell us which distribution to assume. Yet physicists *have* made definite choices, guided by the Principle of Indifference, and they *have* led us to correct and nontrivial predictions of viscosity and many other physical phenomena.
> (1973:478–9)

Another example is the transition from Boltzmann statistics to Bose–Einstein statistics, which is interesting because it is similar to the issue of P versus P^* which arose in connection with Bayesian inference. I will here give a brief informal sketch of the argument.[6] The problem arose with the development of the quantum theory of light by Planck and Einstein. In this theory, cavity radiation becomes analogous to a set of molecules in a gas. Now the problem of the molecules had been worked out using Boltzmann statistics, and it would seem that the same approach could be used for light quanta. However, some modifications were needed to take account of the different situation in the quantum theory. In particular new 'quantum statistics' were introduced by Bose and improved by Einstein. The idea behind these Bose–Einstein statistics is that, although the classical particles used to calculate the Boltzmann statistics were assumed to be distinguishable, light quanta should be regarded as indistinguishable. Consider two particles a, b, say, and suppose each particle either can or cannot have some property M. Let us write 1 if the particle has M, and 0 if it does not. Then, in the classical case, if a is written before b, we have four possible situations 00, 01, 10, 11 which in the Boltzmann statistics are assigned equal probabilities. If, however, the particles are indistinguishable, then we cannot distinguish between 01 and 10, which collapse into a single case. We thus have three possibilities which in the Bose–Einstein statistics are given equal probabilities. The Boltzmann statistics and Bose–Einstein statistics are thus related in more or less the same way as P and P^* in the Bayesian inference case. There is, however, an important difference between the two cases. In the Bayesian inference case, P^* was preferred to P for essentially *ad hoc* reasons, because that choice produced satisfactory results as regards learning from experience. In the physics case, however, there was a good reason in the analysis of cavity radiation with the quantum theory of light to change from Boltzmann to

Bose–Einstein statistics, namely the argument from the indistinguishability of the light quanta.

Let us now return to Jaynes's defence of the Principle of Indifference. He is undoubtedly right to say that this principle has been successfully applied in physics. However, this seems to me to show the fruitfulness of the principle as a *heuristic* principle *not* its validity as a *logical* principle. The Principle of Indifference together with additional considerations such as invariance requirements, arguments about the distinguishability or otherwise of particles, etc. has been, and perhaps will be in the future, very useful for suggesting hypotheses in physics, but the principle did not establish the truth of these hypotheses. They had to be tested empirically like any other hypotheses in physics, and could only be accepted if the predictions derived from them agreed with observation. The heuristic successes of the Principle of Indifference in no way establish that it is a logical principle capable of showing hypotheses to be correct independently of experience.

This point of view can be further illustrated by considering again Jaynes's analysis of the random chord case and the confirming experiment he performed. The same conclusion could have been reached by another scientist (Mr K say) following a different route. Let us suppose that Mr K applies the Principle of Indifference to the random chord case, but initially only the third approach (which yields P(CLSE) = $^1/_4$) occurs to his mind. He works out the full distribution of chord lengths on this approach and then tests the distribution using exactly the same experiment as Jaynes. In the case of Mr K, however, the experiment disproves his hypothetical distribution. In the face of this refutation, Mr K analyses the problem further. He hits on the other two ways of applying the Principle of Indifference, and he also thinks of the invariance requirements which suggest that the first approach is the best of the three. In this way he explains the result of his experiment successfully. The Principle of Indifference was just as valuable a heuristic tool for Mr K as for Jaynes, even though it initially led him to the wrong result. Heuristic principles do not have to give the correct answer every time in order to be fruitful. They have to suggest hypotheses whose testing (and perhaps refutation) will lead to progress. The Principle of Indifference seems to me to have these qualities of a heuristic, but not those of a logical, principle which can be used to demonstrate a result.

The logical interpretation of probability does, however, require the Principle of Indifference to be a logical principle. Only if the Principle of Indifference is logical in character can the logical interpretation allow numerical probabilities. Moreover, as already pointed out, an interpretation of probability which does not allow numerical probabilities can hardly be said to be adequate. Thus the failure to provide a satisfactory solution to the paradoxes of the Principle of Indifference seems to me to be fatal to the logical theory of probability. This point can be reinforced by comparing the situation here with that in deductive logic following the discovery of Russell's paradox.

Russell's paradox follows very simply from a principle known as the axiom of comprehension. Before the discovery of the paradox, this principle had been assumed by the leading logicians, in particular by Frege, Dedekind and Peano (for citations, see Gillies 1982:92). After the paradox had emerged, attempts were made to replace the axiom of comprehension by other principles which would not lead to any contradictions. Russell introduced the theory of types for this purpose. Zermelo, who had discovered the paradox independently, introduced a system of axiomatic set theory which was later improved by Skolem and Fraenkel. Another system of axiomatic set theory was developed by Von Neumann, Bernays and Gödel. Among mathematicians the system stemming from Zermelo is perhaps the most popular, although computer scientists often use the theory of types. In a sense, however, all three systems have been successful. Although it is not possible to prove that they are consistent, they do successfully block the derivation of all known logical paradoxes, and, in more than sixty years of operating in these systems no further contradictions have appeared. This situation contrasts very sharply with that regarding the Principle of Indifference. There is no clearly formulated modification of that principle which blocks the derivation of all the paradoxes. On the contrary, modifications which block some of the paradoxes allow others. Altogether there seems at present to be little hope of successfully rehabilitating the Principle of Indifference as a logical principle.

Keynes uses his Moorean theory of Platonic intuition to ground the logical interpretation, and, in particular, to justify the axioms from this point of view. However, as we shall see in the next chapter, this theory of Platonic intuition is also liable to very grave objections. Altogether the difficulties in the logical interpretation had, by the 1920s, reached such a level that the Bayesians really needed a new interpretation of probability if they were to continue to be able to defend their position. This new interpretation, the subjective theory of probability, did, however, emerge, and I will consider it in the next chapter.

4 The subjective theory

> So have I heard and do in part believe it.
>
> (Shakespeare, *Hamlet*: I, i, 166)

The subjective theory of probability was discovered independently and at about the same time by Frank Ramsey in Cambridge and Bruno de Finetti in Italy. Such simultaneous discoveries are not in fact uncommon in the history of science and mathematics. Usually, however, although the independent discoverers share a common set of ideas, their treatments of the subject differ both in details and in general approach. These differences are of considerable interest, since they illustrate some of the possible variations in the theory. A detailed comparison of the views of Ramsey and De Finetti has recently been published by Galavotti (1989, 1991, 1994) in an important series of papers. In the course of expounding the subjective theory, I will discuss at various points some of these differences between Ramsey and De Finetti.

The existence of simultaneous discoveries is not perhaps so surprising. Usually there is a problem situation in the subject, and the discoverers react to this by producing somewhat similar solutions. We have seen in the previous chapter that by the mid-1920s there were many severe problems in the tradition of logical Bayesian which went back to Bayes and Laplace. Some statisticians (notably Fisher and Neyman) and some philosophers of science (notably Popper) reacted to this by rejecting Bayesianism altogether. However, another approach was to devise a new version of Bayesianism which overcame the difficulties of logical Bayesianism. This was what Ramsey and De Finetti achieved with their new subjective approach to probability.

Since Ramsey's key paper is usually referred to as Ramsey (1926) and De Finetti's earliest publications have later dates, it may appear that Ramsey is the first discoverer and that De Finetti hit on the same idea rather later. This impression is somewhat misleading, however. Ramsey's paper 'Truth and Probability' was written in 1926, and a large part of it read to the Moral Sciences Club at Cambridge, but it was not actually published until 1931. Ramsey died at the age of only 26 in 1930, having made major contributions to the foundations of mathematics, the philosophy of probability, mathematical logic and economics. His paper on probability first appeared in the collection published after his early death in 1931. De Finetti says that already by April 1928 he had written a complete exposition of the foundations of probability theory according to the subjective point of view. This may have been a little later than Ramsey, but De Finetti was the first to publish (1930a, b, c). In 1931 De Finetti (1931a) gave a full account of the philosophical aspects of the theory without formulas in his ‘Probabilism’, and provided more details about the mathematical foundations in his 1931b paper. Ramsey certainly never heard of De Finetti, and De Finetti seems not to have read Ramsey until after 1937, when his own views had been completely developed [see his new footnote (a) added in 1964 to 1937:102]. Thus, the discovery was completely independent and occurred at almost the same time.

Ramsey’s relation to the older logical tradition is very clear, since he introduces his new theory by giving detailed criticisms of Keynes’s views. De Finetti, however, does not appear to have been influenced by Keynes at the time when he devised the subjective theory. Indeed in his 1931a paper, he seems to be doubtful about what exactly Keynes’s views were, remarking in a footnote: ‘This seems to me to be Keynes’s point of view; but I cannot judge well, since I have only been able to skim his essay quickly.’ (1931a:221). Later, De Finetti expounds and criticises Keynes’s views, and remarks in a footnote: ‘I briefly saw Keynes’s book in 1929 (and I quoted it in ‘Probabilismo’ ... 1931 ...), understanding little of it, however, because of my then insufficient knowledge of English. This year I have read the German version’ (1938:362, Footnote 18). It thus seems clear that De Finetti properly studied Keynes only after his own views had been fully developed. It is also interesting to note that De Finetti’s 1938 paper is entitled ‘Cambridge Probability Theorists’; he mentions only Keynes and Jeffreys, but not Ramsey. This indicates that he probably only read Ramsey after 1938. In the light of all this, I will begin the next section with Ramsey’s criticisms of Keynes, since these follow on naturally from the previous chapter. However in the section ‘Some objections to Bayesianism’ I will give some consideration to De Finetti’s different route to subjective probability. The remaining sections will expound the subjective theory itself. ‘Subjective foundations for mathematical probability’ shows how the mathematical theory of probability can be developed on the subjective approach, and, in particular, gives a full proof of the all important Ramsey–De Finetti theorem. ‘Apparently objective probabilities in the subjective theory’ introduces the key notion of *exchangeablility,* which, as we shall see, plays a most important rôle in the theory. Both these sections are largely based on De Finetti (1937), which is my own preferred account of the theory. However, I will introduce a few changes and amplifications for the sake of clarity and will also mention some alternatives to be found in Ramsey and in De Finetti’s later work. ‘A comparison of the axiom system given here with the Kolmogorov axioms*’ and ‘The relation between independence and exchangeability*’ cover some rather mathematical points, and in another section I will present my criticism of De Finetti’s exchangeability reduction.

Ramsey's criticisms of Keynes[1]

According to Keynes there are logical relations of probability between pairs of propositions, and these can be in some sense perceived. Ramsey criticises this as follows:

> But let us now return to a more fundamental criticism of Mr. Keynes' views, which is the obvious one that there really do not seem to be any such things as the probability relations he describes. He supposes that, at any rate in certain cases, they can be perceived; but speaking for myself I feel confident that this is not true. I do not perceive them, and if I am to be persuaded that they exist it must be by argument; moreover I shrewdly suspect that others do not perceive them either, because they are able to come to so very little agreement as to which of them relates any two given propositions.
>
> (1926:161)

This is an interesting case of an argument which gains in strength from the nature of the person who proposes it. Had a less distinguished logician than Ramsey objected that he was unable to perceive any logical relations of probability, Keynes might have replied that this was merely a sign of logical incompetence, or logical blindness. Indeed Keynes does say: 'Some men – indeed it is obviously the case – may have a greater power of logical intuition than others.' (1921:18). Ramsey, however, was not just a brilliant mathematical logician but a member of the Cambridge Apostles as well. Thus Keynes could not have claimed with plausibility that Ramsey was lacking in the capacity for logical intuition or perception – and Keynes did not in fact do so.

Ramsey buttresses his basic argument by pointing out that, on the logical theory, we can apparently perceive logical relations in quite complicated cases, while being quite unable to perceive them in simple cases. Thus he says:

> All we appear to know about them [i.e. Keynes's logical relations of probability] are certain general propositions, the laws of addition and multiplication; it is as if everyone knew the laws of geometry but no one could tell whether any given object were round or square; and I find it hard to imagine how so large a body of general knowledge can be combined with so slender a stock of particular facts. It is true that about some particular cases there is agreement, but these somehow paradoxically are always immensely complicated; we all agree that the probability of a coin coming down heads is 1/2, but we can none of us say exactly what is the evidence which forms the other term for the probability relation about which we are then judging. If, on the other hand, we take the simplest possible pairs of propositions such as 'This is red' and 'That is blue' or 'This is red' and 'That is red', whose logical relations should surely be easiest to see, no one, I think, pretends to be sure what is the probability relation which connects them.
>
> (Ramsey 1926:162)

Ramsey's doubts about basing probability theory on logical intuition are reinforced by considering how logical intuition fared in the case of deductive inference, which is surely less problematic than inductive. Frege, one of the greatest logicians of all time, was led by his logical intuition to support the so-called axiom of comprehension, from which Russell's paradox follows in a few lines. Moreover, he had companions in this error as distinguished as Dedekind and Peano (for citations, see Gillies 1982: 92). Hilbert and Brouwer were two of the greatest mathematicians of the twentieth century. Yet Hilbert's logical intuition informed him that the Law of the Excluded Middle was valid in mathematics, and Brouwer's that it was not valid. All this indicates that logical intuition is not to be greatly trusted in the deductive case, and so hardly at all as regards inductive inferences.

Moreover, is so-called logical intuition anything more than a psychological illusion caused by familiarity? Perhaps it is only as a result of studying the mathematical theory of probability for several years that the axioms come to seem intuitively obvious. Maybe the basic principles of Aristotle's philosophy seemed intuitively obvious to scholars in medieval Europe, and those of Confucian philosophy to scholars in China at the same time. I conclude that logical intuition is not adequate to establish either that degrees of partial entailment exist, or that they obey the usual axioms of probability. Let us accordingly examine in the next section how these matters are dealt with in the subjective theory.

Subjective foundations for mathematical probability: the Ramsey–De Finetti theorem

In the logical interpretation, the probability of h given e is identified with the rational degree of belief which someone who had evidence e would accord to h. This rational degree of belief is considered to be the same for all rational individuals. The subjective interpretation of probability abandons the assumption of rationality leading to consensus. According to the subjective theory, different individuals (Ms A, Mr B and Master C say), although all perfectly reasonable and having the same evidence e, may yet have different degrees of belief in h. Probability is thus defined as the degree of belief of a particular individual, so that we should really not speak of *the* probability, but rather of Ms A's probability, Mr B's probability or Master C's probability.

Now the mathematical theory of probability takes probabilities to be numbers in the interval [0, 1]. So, if the subjective theory is to be an adequate interpretation of the mathematical calculus, a way must be found of measuring the degree of belief of an individual that some event (E say) will occur. Thus, we want to be able to measure, for example, Mr B's degree of belief that it will rain tomorrow in London, that a particular political party will win the next election, and so on. How can this be done?

Ramsey has an interesting discussion of this problem. His first remark on the question is that 'it is, I suppose, conceivable that degrees of belief could be measured by a psychogalvanometer or some such instrument' (1926:161). Ramsey's psychogalvanometer would perhaps be a piece of electronic apparatus something like a superior lie detector. We would attach the electrodes to Mr B's skull, and, when he read out a proposition describing the event E in question, the machine would register his degree of belief in that proposition. Needless to say, even if such a psychogalvanometer is possible at all, no such machine exists at present, and we cannot solve our problem of measuring belief in this way.

Ramsey next considers the possibility of using introspection to estimate the strength of our belief-feeling about some proposition. However, he has an interesting argument against such an approach:

> We can, in the first place, suppose that the degree of a belief is something perceptible by its owner; for instance that beliefs differ in the intensity of a feeling by which they are accompanied, which might be called a belief-feeling or feeling of conviction, and that by the degree of belief we mean the intensity of this feeling. This view would be very inconvenient, for it is not easy to ascribe numbers to the intensities of feelings; but apart from this it seems to me observably false, for the beliefs which we hold most strongly are often accompanied by practically no feeling at all; no one feels strongly about things he takes for granted.
>
> (1926:169)

Ramsey is undoubtedly correct here. When I cut a slice of bread to eat, I believe very strongly that it will nourish rather than poison me, but this belief, under normal circumstances, is not accompanied by any strong feelings, or indeed any feelings at all. Ramsey is thus led to the conclusion that: '... the degree of a belief is a causal property of it, which we can express vaguely as the extent to which we are prepared to act on it' (1926:169). I am certainly prepared to act on my belief that the bread is nourishing rather than poisonous by eating it without hesitation, even though I am not having any strong feelings at the time.

On this approach we should measure the strength of a belief by examining the character of some action to which it leads. A suitable action for measurement purposes is betting, and so Ramsey concludes: 'The old-established way of measuring a person's belief is to propose a bet, and see what are the lowest odds which he will accept. This method I regard as fundamentally sound' (1926:172). De Finetti (1930a) also introduces bets to measure degrees of belief.

Betting is of course just one kind of action to which a belief can lead. Does it therefore give a good measure of the strength of a belief as regards other sorts of actions to which a belief might lead? Ramsey defends the assumption that it does as follows:

> ... this section ... is based fundamentally on betting, but this will not seem unreasonable when it is seen that all our lives we are in a sense betting. Whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home.
>
> (1926:183)

My own view is that betting does give a reasonable measure of the strength of a belief in many cases, but not in all. In particular, betting cannot be used to measure the strength of someone's belief in a universal scientific law or theory (for a discussion, see Gillies 1988a:192–5). However, let us for the moment accept betting as a reasonable way of measuring degree of belief and see what this assumption leads to.

To do this, we must now present some mathematics, but, since the purpose of this book is to discuss the philosophical aspects of probability, I have tried to keep this mathematics as simple as possible, and indeed it involves no more than elementary algebra. We must first set up a hypothetical betting situation in which the rate at which Mr B is prepared to bet on E (his *betting quotient* on E) can be taken as a measure of his degree of belief in E. Then we introduce the condition of *coherence.* It will be clear that Mr B ought to choose his betting quotients in order to be coherent, and this leads to the main result (*The Ramsey–De Finetti Theorem*), which states that a set of betting quotients is coherent if and only if they satisfy the axioms of probability. I will state the axioms of probability in full and then prove the Ramsey–De Finetti theory for each one. In this way the foundations of the mathematical theory of probability will be established from the subjective point of view.

Definition of betting quotients (q)

We imagine that Ms A (a psychologist) wants to measure the degree of belief of Mr B in some event E.[2] To do so, she gets Mr B to agree to bet with her on E under the following conditions. Mr B has to choose a number q (called his *betting quotient* on E), and then Ms A chooses the stake S. Mr B pays Ms A qS in exchange for S if E occurs. S can be positive or negative, but $|S|$ must be small in relation to Mr B's wealth. Under these circumstances, q is taken to be a measure of Mr B's degree of belief in E.

A number of comments on this definition are in order. First of all it is important that Mr B does not know when choosing q whether the stake S will be positive (corresponding to his betting in favour of the event E occurring) or whether S will be negative (corresponding to his betting against E). If Mr B knew that S would be positive, it would be in his interest to choose q as low as possible. If he knew S would be negative, it would be in his interest to choose q as high as possible. In neither case would q correspond to his true degree of belief. However, if he does not know whether S is going to be positive or negative, he has to adjust q to his actual belief.

We can illustrate this by a hypothetical example from the stock market. Suppose Mr B is now a jobber, and I want to find out what he thinks to be the value of a particular share (BP say). If I say to him: 'I want to sell 100 BP shares, what do you think their value is?', it will be in Mr B's interest to quote a value rather below what he thinks to be the correct one, since in this way he can hope to pick up some BP shares cheaply. Conversely, if I say to him:'I want to buy 100 BP shares, what do you think their value is?', it will be in Mr B's interest to quote a value rather above what he thinks to be the correct one, since in this way he can hope to sell some BP shares at a good profit. If, however, I ask Mr B's opinion as to the value of a BP share without saying whether I want to buy or sell, he will be forced to state his true opinion as to the value. Of course, this is only a hypothetical example to illustrate the point. In actual stock market practice, jobbers quote one price for buying and one for selling.

My next point concerns the way in which the magnitude of the stake *S* is measured, for here there is a difference between De Finetti (at least in his early papers) and Ramsey. De Finetti took the stakes to be in money, whereas Ramsey developed a theory of utility and took the stakes to be in utility as he had defined it. My own preference is for De Finetti's early approach, i.e. stakes in money, and I will now briefly discuss some of the issues involved.

If the bets are to be in money, then it is obvious that the sums used should not be too large – at least in relation to Mr B's fortune. Suppose Mr B's entire savings amount to £500. Then it would not be reasonable for Ms A to propose a bet with him on whether it will rain tomorrow with a stake of £500. On the other hand, if Mr B happens to be a billionaire, a stake of £500 might not be unreasonable, provided Ms A's research grant can cover bets of this magnitude.

Ramsey thinks that difficulties of this sort constitute a serious objection to money bets, for he writes: '... if money bets are to be used, it is evident that they should be for as small stakes as possible. But then again the measurement is spoiled by introducing the new factor of reluctance to bother about trifles.' (1926:176). It seems to me, however, that this difficulty can be overcome. Ms A has to choose a size of stake which is small enough in relation to Mr B's fortune so that the bet will not damage him financially but which is large enough to make him think seriously about the bet. I think that it would, in general, be possible to find such a level for the stakes, especially as we have to imagine Mr B as co-operating with the psychological experiment of trying to measure his degree of belief. If Mr B were totally averse to such an experiment, it would hardly be possible to carry it out.

Although there do not seem to me any major objections to money bets, I regard the introduction of a satisfactory measure of utility as a virtually impossible task. We can see some of the difficulties by giving a few quotations which illustrate Ramsey's own procedure. Ramsey writes:

> Let us call the things a person ultimately desires 'goods', and let us at first assume that they are numerically measurable and additive. That is to say that if he prefers for its own sake an hour's swimming to an hour's reading, he will prefer two hours' swimming to one hour's swimming and one hour's reading. This is of course absurd in the given case but this may only be because swimming and reading are not ultimate goods, and because we cannot imagine a second hour's swimming precisely similar to the first, owing to fatigue, etc.
>
> (1926:173–4)

I find it hard to believe that there is any satisfactory way of comparing the utility of an hour's swimming with that of an hour's reading. Both can give considerable pleasure, but the pleasures are of quite a different kind and so incomparable. Ramsey thinks that this difficulty can be overcome by introducing 'ultimate goods'. But what are these ultimate goods? No ultimate good is ever specified, and such a thing would appear to be a myth rather than a reality.

At another stage of his introduction of utility, Ramsey writes: '... we could, by offering him options, discover how he placed in order of merit all possible courses of the world. In this way all possible worlds would be put in an order of value' (1926:176). Such a procedure seems to belong to the realm of pure fantasy. Compare it with the realistic possibility of betting for a stake of £1 on whether it will rain tomorrow.

It might be objected that these arguments are directed just against Ramsey's way of introducing measurable utility, and that other more satisfactory methods might be available. Yet other methods involve similar difficulties and often lead to curious paradoxes which are difficult to resolve. Surely it is better to avoid this minefield and just consider money bets made with appropriate stakes. This latter procedure, far from belonging to the realm of fantasy can easily be carried out in practice. Indeed, De Finetti used to get his class of students to produce betting quotients on the results of Italian football games. Being of a democratic turn of mind, he invited the porter to participate as well, and the porter was nearly always the most successful. He knew more than anyone else about football.

A further objection to the betting scheme might be that it produces only very rough estimates and hardly exact numerical probabilities. De Finetti's reply to this point is that exact numerical degrees of belief are indeed something of a fiction or idealisation, but that this idealisation is a useful one in that it simplifies the mathematical calculations. Moreover, provided we do not forget that the mathematics must be understood as holding approximately, this idealisation does no harm. As De Finetti himself says:

> ... if you want to apply mathematics, you must act as though the measured magnitudes have precise values. This fiction is very fruitful, as everybody knows; the fact that it is only a fiction does not diminish its value as long as we bear in mind that the precision of the result will be what it will be.... To go, with the valid help of mathematics, from approximate premises to approximate conclusions, I must go by way of an exact algorithm, even though I consider it an artifice.
>
> (1931a:204)

My own conclusion then is that we should use the betting scheme with money bets and appropriately selected stakes, and that this does indeed give a reasonable method for measuring belief in many situations. I therefore adhere to the approach of the early De Finetti. Curiously, however, De Finetti in his later period moved in the direction of using utility, and in his last papers even abandoned the betting approach altogether. In 1957 De Finetti still hesitated to follow Savage in trying to unify probability and utility within decision theory (see quotation in Galavotti 1989:240). However, in 1964 in a new footnote to his 1937 paper he wrote: 'Such a formulation could better, like Ramsey's, deal with expected *utilities*' (p. 102). In his 1970 book he used mainly decision theory to introduce subjective probabilities. He also develops a theory of utility, even though he still seems to regard this with some degree of scepticism (see De Finetti 1970:76–82). In one of his very last papers, he went as far as to repudiate the whole betting approach as inadequate, writing: '... betting, strictly speaking, does not pertain to probability but to the Theory of Games ... It is because of this that I invented and applied in experiments (probabilistic forecasts) the "proper scoring rules"' (De Finetti 1981b:55). Thus, De Finetti himself moved in the direction of decision theory and utilities. However, for reasons already given, my own preference is for De Finetti's earlier approach, and this is what I will use as the basis of the account which follows.[3]

The first problem in the subjective approach was how to measure degrees of belief. We have seen how the betting scheme offers a reasonable solution to this problem. Mr B's degree of belief in E is measured by his betting quotient in E as elicited in the situation described above. It is worth noting that this way of introducing probabilities is in accordance with the philosophy of *operationalism.* A recent important contribution to subjective probability is Lad (1996). In this book, Lad provides a foundation for subjective probability similar to De Finetti's but goes beyond De Finetti by showing in detail how statistics can be developed from this point of view. In the title of his book and throughout the book itself, Lad speaks of 'operational subjective statistical methods', which emphasises the point that subjective probability is based on operationalism. Lad writes: 'An *operationally defined measurement* is a specified procedure of action which, when followed, yields a number.' (1996:39). It is clear that the measurement of degrees of belief by betting quotients as just described is an operationally defined measurement in this sense. We shall return to this connection between subjective probability and operationalism from time to time in what follows.

Let us now examine a second problem which arises in the subjective approach. If the subjective theory is to provide an interpretation of the standard mathematical theory of probability, then these degrees of belief (or betting quotients) ought to satisfy the standard axioms of probability. But why should they do so? It seems easy to imagine an individual whose degrees of belief are quite arbitrary and do not satisfy any of the axioms of probability. The subjectivists solve this problem and derive the axioms of probability by using the concept of *coherence.* I will next define this concept and then comment on its significance.

Coherence

If Mr B has to bet on a number of events $E_1, ..., E_n$, his betting quotients are said to be *coherent* if and only if Ms A cannot choose stakes $S_1, ..., S_n$ such that she wins whatever happens. If Ms A can choose stakes so that she wins whatever happens, she is said to have made a *Dutch book* against Mr B.

It is taken as obvious that Mr B will want his bets to be coherent, that is to say he will want to avoid the possibility of his losing whatever happens. Surprisingly, this condition is both necessary and sufficient for betting quotients to satisfy the axioms of probability. This is the content of the following theorem.

The Ramsey–De Finetti theorem

A set of betting quotients is coherent if and only if they satisfy the axioms of probability.

So far we have made a contrast between the logical theory, in which probability is degree of rational belief, and the subjective theory, in which probability is degree of belief. The concept of coherence shows that this needs a little qualification, since coherence is after all a rationality constraint, and degrees of belief in the subjective approach must be rational, at least to the extent of satisfying this constraint. De Finetti expresses this very well in the title of his 1937 paper 'Foresight: Its Logical Laws, Its Subjective Sources'. The logical laws here come from the condition of coherence. Naturally, coherence does not determine a single degree of rational belief but leaves open a wide range of choices. Thus some subjective sources for probability are also needed.

Ramsey uses the term 'consistency' for coherence, and writes that: '... the laws of probability are laws of consistency' (1926:182). The idea here is that we have to make sure that our various degrees of belief fit together and so avoid the 'contradiction' of having a Dutch book made against us. The term 'coherence' is now generally preferred, because consistency has a well-defined but different meaning in deductive logic. Even though there is an analogy, it seems better to use different terms. I will now give a detailed proof of the Ramsey–De Finetti theorem. First I will state the axioms of probability and then prove the theorem for each of them in turn.

The axioms of probability

Let E, F, ..., E_1, ... stand for events, concerning which we can have some degree of belief whether they will occur, or have occurred. Let Ω denote the certain event, which must occur. There are then three axioms of probability.

1 $0 \leq P(E) \leq 1$ for any E, and $P(\Omega) = 1$.

2 (Addition Law) If $E_1, \ldots, E_n$ are events which are exclusive (i.e. no two can both occur) and exhaustive (i.e. at least one must occur), then

$$P(E_1) + \ldots + P(E_n) = 1$$

3 (Multiplication Law) For any two events E, F

$$P(E \,\&\, F) = P(E \mid F)\, P(F)$$

The Addition Law can be stated in a different but equivalent form. For any event E, F, let E v F be the event that either E occurs or F occurs or both occur. Then we have

2' (Alternative form of the Addition Law) If E, F are any two exclusive events, then

$$P(E) + P(F) = P(E \vee F)$$

We can prove the equivalence of 2 and 2' as follows:

(a) (2 → 2') Let E, F be exclusive events, and let $\Omega \setminus (E \vee F)$ be the event that something other than E or F occurs. E, F, $\Omega \setminus (E \vee F)$ are exclusive and exhaustive events. So by Axiom 2

$$P(E) + P(F) + P(\Omega \setminus (E \vee F)) = 1$$

But E v F, $\Omega \setminus (E \vee F)$ are also exclusive and exhaustive events. So by Axiom 2

$$P(E \vee F) + P(\Omega \setminus (E \vee F)) = 1$$

Thus subtracting, we get

$P(E) + P(F) = P(E \vee F)$ i.e. Axiom 2'

(b) (2' → 2) We first prove by induction that Axiom 2' holds for any n exclusive events. The case $n = 2$ is just Axiom 2' itself. Suppose the result holds for $n - 1$, i.e. if $E_1, ..., E_{n-1}$ are any exclusive events, then

$$P(E_1) + ... + P(E_{n-1}) = P(E_1 \vee ... \vee E_{n-1})$$

Now consider n exclusive events $E_1, ..., E_n$. The events $(E_1 \vee ... \vee E_{n-1})$, E_n are also exclusive. So by Axiom 2'

$$P(E_1 \vee ... \vee E_{n-1}) + P(E_n) = P(E_1 \vee ... \vee E_n)$$

But since $E_1, ..., E_{n-1}$ are exclusive events, it follows that

$$P(E_1) + ... + P(E_n) = P(E_1 \vee ... \vee E_n)$$

But if $E_1, ..., E_n$ are exhaustive as well as exclusive, $E_1 \vee ... \vee E_n$ is the certain event with probability 1, and so Axiom 2 follows.

Proof of the Ramsey–De Finetti theorem [4]

Proof for Axiom 1

(a) Coherence → Axiom 1: Let us first consider the case of the certain event O. If Mr B chooses $q(\Omega) > 1$, Ms A can win by choosing $S > 0$. If Mr B chooses

$q(\Omega) < 1$, Ms A can win by choosing $S < 0$. Hence to be coherent, Mr B must choose $q(\Omega) = 1$. Now take any arbitrary event E. If Mr B chooses $q(\mathrm{E}) > 1$, Ms A can win by choosing $S > 0$. If Mr B chooses $q(\mathrm{E}) < 0$, Ms A can win by choosing $S < 0$. Hence to be coherent, Mr B must choose $0 = q(\mathrm{E}) = 1$.

(b) Axiom 1 → coherence: If Mr B chooses $q(\Omega) = 1$, there is no way that Ms A can win, since the stake, whatever its sign, is simply passed from one to the other and then back again. For an arbitrary event E, Ms A cannot choose the sign or size of S so that she always wins if Mr B chooses $0 = q(\mathrm{E}) = 1$.

Proof for Axiom 2

(a) Coherence → Axiom 2: Suppose Mr B chooses betting quotients $q_1, \ldots, q_n$, and Ms A chooses stakes $S_1, \ldots, S_n$. Then, if event E_i occurs, Ms A's gain G_i is given by

$$G_i = q_1S_1 + \ldots + q_nS_n - S_i \tag{4.1}$$

So if Ms A sets $S_1 = S_2 = \ldots = S_n = S$, then

$$G_i = S(q_1 + \ldots + q_n - 1)$$

Thus, if Mr B chooses $q_1 + \ldots + q_n > 1$, then Ms A can always win by setting $S > 0$. If Mr B chooses $q_1 + \ldots + q_n < 1$, then Ms A can always win by setting $S < 0$. Hence, to be coherent, Mr B must choose $q_1 + \ldots + q_n = 1$.

(b) Axiom 2 → coherence: Since Axiom 2 holds, we have $q_1 + \ldots + qn = 1$. Now by Equation 4.1 above, we have

$$q_iG_i = q_i(q_1S_1 + \ldots + q_nS_n) - q_iS_i$$

So summing over *i*, we get

$$q_1G_1 + q_2G_2 + \ldots + q_nG_n = 0 \tag{4.2}$$

Equation 4.2 shows that the Gi cannot all be positive for the following reason. The $q_i = 0$, and, since they sum to 1, at least one of them must be > 0. Hence if all the G_i were > 0, $q_1G_1 + \ldots + q_nG_n > 0$, which contradicts Equation 4.2. Hence, not all the Gi can be positive, which is equivalent to saying that the betting quotients are coherent. The consideration of $q_1G_1 + q_2G_2 + \ldots + q_nG_n$ may look like a mathematical trick, but in fact it has a simple intuitive meaning.[5] It is just Ms A's expected gain relative to the probabilities chosen by Mr B. If this expected gain is zero, Ms A cannot make a Dutch book against Mr B.

To prove the Ramsey–De Finetti theorem for Axiom 3, we need the following definition.

Definition of conditional betting quotient

$q(\mathrm{E} \mid \mathrm{F})$, the conditional betting quotient for E given F, is the betting quotient which Mr B would give for E on the understanding that the bet is called off and all stakes returned if F does not occur.

Ramsey remarks that 'Such conditional bets were often made in the eighteenth century.' (1926:180).

Proof for Axiom 3

In all parts of the proof, we shall use the following notation

$$q = q(\mathrm{E}\ \&\ \mathrm{F})$$
$$q' = q(\mathrm{E} \mid \mathrm{F})$$
$$q'' = q(\mathrm{F})$$

(a) Coherence → Axiom 3, using determinants: Suppose Mr B chooses betting quotients q, q', q'' as above, and Ms A chooses corresponding stakes S, S', S''. Three possible cases can occur, and we shall calculate Ms A's gain in each case.

1 E and F both occur

$$G_1 = (q - 1)\,S + (q' - 1)S' + (q'' - 1)\,S''$$

2 E does not occur, but F occurs

$$G_2 = qS + q'S' + (q'' - 1)\,S''$$

3 F does not occur

$$G_3 = qS + \quad + q''S''$$

For fixed $G_1, G_2, G_3 > 0$, these are three linear equations in three unknowns, S, S', S''. Thus, they always have a solution, unless the determinant vanishes. So, for coherence, we must have

$$\begin{vmatrix} q-1 & q'-1 & q''-1 \\ q & q' & q''-1 \\ q & 0 & q'' \end{vmatrix} = 0$$

Subtracting the bottom row from the top two rows, and then the middle row from the top row gives

$$\begin{vmatrix} -1 & -1 & 0 \\ 0 & q' & -1 \\ q & 0 & q'' \end{vmatrix} = 0$$

Then expanding by the first row, we get

$$- q'q'' + q = 0$$
So $q = q'q''$ as required.

For those unfamiliar with the theory of determinants, the following gives a proof of the same result without using determinants.

(b) Coherence $\rightarrow$ Axiom 3, without using determinants: Suppose Ms A chooses $S = +1$, $S' = -1$, $S'' = -q'$, we then have

$$G_1 = (q - 1) + (1 - q') + q' - q'q'' = q - q'q''$$

$$G_2 = q - q' - q'q'' + q' = q - q'q''$$

$$G_3 = q - q'q''$$

So all Ms A's gains are positive, unless $q \leq q'q''$.

Similarly, if Ms A chooses $S = -1$, $S' = +1$, $S'' = q'$, all her gains are positive unless $q \geq q'q''$. So, to be coherent, Mr B must choose $q = q'q''$, as required.

(c) Axiom 3 $\rightarrow$ coherence: We have to show that if $q = q'q''$, the betting quotients are coherent, i.e. Ms A's gains G_1, G_2, G_3 cannot all be positive. Using the method employed for Axiom 2, we need to consider Ms A's expected gain given the probabilities chosen by Mr B, and then show that it is zero. Ms A's expected gain is in fact $\lambda_1 G_1 + \lambda_2 G_2 + \lambda_3 G_3$ where

$\lambda_1 = q'q''$, $\lambda_2 = (1 - q')q''$, $\lambda_3 = 1 - q''$. Since $0 \leq q', q'' \leq 1$, each $\lambda_i \geq 0$.

Now

$$\lambda_1 G_1 + \lambda_2 G_2 + \lambda_3 G_3 = \alpha S + \beta S' + \gamma S'',$$

where

$$\alpha = q'q''(q - 1) + (1 - q')q''q + (1 - q'')\, q$$

$$= q''(q'q - q' + q - qq' + (1 - q'')q'), \text{ since } q = q'q''$$

$$= q''(q'q - q' + q'q'' - qq' + q' - q'q'')$$

$$= 0$$

$\beta = q'q''(q' - 1) + (1 - q')q''q' = 0$

$\gamma = q'q''(q'' - 1) + (1 - q')q''(q'' - 1) + (1 - q'')q'' = 0$

Hence $\lambda_1 G_1 + \lambda_2 G_2 + \lambda_3 G_3 = 0$.

But now at least one of the $\lambda_i > 0$, for either $q'' \neq 1$, when $\lambda_3 > 0$, or $q'' = 1$, when $\lambda_1 = q'$, $\lambda_2 = 1 - q'$. In this case, either $q' \neq 1$, when $\lambda_2 > 0$, or $q' = 1$, when $\lambda_1 > 0$. It follows that not all the G_i can be positive, and so Mr B's betting quotients are coherent, as required.

The Ramsey–De Finetti theorem is a remarkable achievement, and clearly demonstrates the superiority of the subjective to the logical theory. Whereas in the logical theory the axioms of probability could only be justified by a vague and unsatisfactory appeal to intuition, in the subjective theory they can be proved rigorously from the eminently plausible condition of coherence. Indeed, given the Ramsey–De Finetti theorem, it is difficult to deny that the subjective theory provides a valid interpretation of the mathematical calculus of probability – though it is of course possible to hold that there are other valid interpretations of this calculus. In addition, the subjective theory solves the paradoxes of the Principle of Indifference by, in effect, making this principle unnecessary, or at most a heuristic device. In the logical theory, the principle was necessary to obtain the supposedly unique a priori degrees of rational belief, but, according to the subjective theory, there are no unique a priori probabilities. Different individuals can choose their a priori probabilities in different ways, and, provided they are coherent, there need be nothing wrong with these different choices. Thus, if the Principle of Indifference is used as a heuristic device, and suggests two different possibilities for the a priori probabilities, there is no contradiction. Mr B might choose one of these possibilities as his subjective valuation, and Ms D might choose the other. Ramsey is well aware of the superiority of the subjective to the logical theory in these respects and states them as follows:

> In the first place it gives us a clear justification for the axioms of the calculus, which on such a system as Mr Keynes' is entirely wanting. For now it is easily seen that if partial beliefs are consistent they will obey these axioms, but it is utterly obscure why Mr Keynes' mysterious logical relations should obey them. We should be so curiously ignorant of the instances of these relations, and so curiously knowledgeable about their general laws.
>
> Secondly, the Principle of Indifference can now be altogether dispensed with; ... To be able to turn the Principle of Indifference out of formal logic is a great advantage; for it is fairly clearly impossible to lay down purely logical conditions for its validity, as is attempted by Mr Keynes.
>
> (Ramsey 1926:188–9)

There remain, however, some problems connected with the subjective theory, and in particular the question of how probabilities which appear to be objective, such as the probability of a particular isotope of uranium disintegrating in a year, can be explained on this approach. De Finetti tackles this problem by introducing the concept of *exchangeability,* and I will give an account of this below (pp. 69– 83). Before going on to this, however, there is a matter which may well be of interest to mathematicians. Nearly all advanced treatments of mathematical theory of probability are today based on the Kolmogorov axioms (see Kolmogorov 1933). Now the axioms given above are of course similar to the Kolmogorov axioms, but do nonetheless differ on one or two points. It certainly seems worth examining these divergences from standard mathematical practice to see what significance they have. In general, in this book my aim is to discuss the philosophical side of probability using as little mathematics as possible, indeed no more than quite elementary algebra. Sometimes, as here, however, it will be useful to discuss issues which require a knowledge of more advanced mathematical approaches to probability (random variables, measure theory, analysis, etc.). My plan is to place such discussions in sections marked with an asterisk and to arrange them so that they can be read by mathematicians but omitted by non-mathematicians without losing the general thread of the argument.

A comparison of the axiom system given here with the Kolmogorov axioms*

De Finetti assigns probabilities to events E, F, ..., including the certain event which we have denoted by Ω. In Kolmogorov's mathematical approach, probabilities are assigned to the subsets of a set Ω. This difference does not seem to me an important one, since it would be fairly easy to map De Finetti's treatment into set-theoretic language. A more significant divergence comes with the treatment of conditional probabilities. Kolmogorov introduces these by definition (see Kolmogorov 1933:6), so that

$$P(E \mid F) = \text{def}\,\frac{P(E \,\&\, F)}{P(F)} \quad \text{for } P(F) \neq 0$$

The case $P(F) = 0$ is dealt with by Kolmogorov later in his monograph (1933:Chapter V). Thus, in Kolmogorov's treatment an equality is established by definition which in the treatment we have just given is a substantial axiom (Axiom 3) requiring an elaborate proof, and is indeed the multiplication law of probability.

In fact, this is not the only instance in mathematics where a substantial assumption appears in the form of a definition, but the practice does not seem to me a good one. I would argue that it is better to state important assumptions as axioms (or derive them as theorems) and try to keep definitions as far as possible as mere abbreviations. This inclines me to prefer De Finetti's treatment to Kolmogorov's on this point. This would amount to taking $P(E \mid F)$ as a primitive (undefined) term in the axiom system and characterising it by an axiom, rather than introducing it by an explicit definition.

It is clear that De Finetti's approach is more natural for the subjective theory, since conditional probabilities can be introduced as conditional betting quotients defined within a particular betting scheme. It is then by no means obvious that these conditional betting quotients obey our Axiom 3; indeed the proof is quite long. Moreover, similar considerations apply in the other interpretations of probability. We have seen in Chapter 3 that the notion of the conditional probability of h given e is a primitive and fundamental notion within the logical theory. It thus seems natural to take it as a primitive notion in an axiom system, as Keynes does. As we shall see in Chapters 5 and 6, the notion of conditional probability is also primitive in the frequency and propensity interpretations. On this point I side with De Finetti rather than Kolmogorov, and I favour the introduction of conditional probabilities by an axiom rather than a definition. This, moreover, leads to a rather elegant symmetry in the axiomatic treatment between the addition and multiplication laws of probability.

The next important difference between De Finetti and Kolmogorov concerns the issue of finite versus countable additivity. De Finetti's Axiom 2 (the Addition Law) can, as we have seen, be stated in the equivalent form: if $E_1, ..., E_n$ are events which are exclusive,

$$P(E_1 \text{ v } ... \text{ v } E_n) = P(E_1) + ... + P(E_n).$$

The question now arises whether we can extend the Addition Law from the finite case to the countably infinite case, that is to say whether we can legitimately go from finite additivity to countable additivity. This would involve adopting as an axiom the following stronger form of the Addition Law.

> Addition law for countable additivity: If $E_1, ..., E_n, ...$ is a countably infinite sequence of exclusive events, then
>
> $$P(E_1 \text{ v } ... \text{ v } E_n \text{ v } ...) = P(E_1) + ... + P(E_n) + ...$$

Kolmogorov's treatment of this question is interesting. In the first chapter of his monograph he allows only finite additivity. Then in the second chapter he adds to his five previous axioms a sixth axiom (the axiom of continuity) which is equivalent to the Addition Law for countable additivity as just stated. Kolmogorov does, however, appear to have some reservations about his axiom, for he says:

> Since the new axiom is essential for infinite fields of probability only, it is almost impossible to elucidate its empirical meaning, as has been done, for example, in the case of Axioms I – V in §2 of the first chapter. For, in describing any observable random process we can obtain only finite fields of probability. Infinite fields of probability occur only as idealised models of real random processes. *We limit ourselves, arbitrarily, to only those models which satisfy Axiom VI.* This limitation has been found expedient in researches of the most diverse sort.

Kolmogorov here argues that countable additivity goes beyond what can be checked empirically, but that its adoption is nonetheless justified because of its usefulness in a whole range of research.

De Finetti shares Kolmogorov's doubts about countable additivity, but he regards them as a reason for limiting oneself to finite additivity.[6] Thus he says that:

> [The assumption of countable additivity] is the one most commonly accepted at present; it had, if not its origin, its systematization in Kolmogorov's axioms (1933). Its success owes much to the mathematical convenience of making the calculus of probability merely a translation of modern measure theory.... No-one has given a real justification of countable additivity (other than just taking it as a 'natural extension' of finite additivity).
>
> (1970:vol. 1, 119)

De Finetti, however, thinks that one should not introduce new axioms simply on the grounds of mathematical convenience, unless these axioms can be justified in terms of the meaning of probability. Now in the subjective theory, probabilities are given by an individual's betting quotients. A given individual will always bet on a finite number of events, and it is difficult to imagine bets on an infinite number of events. Thus the subjective theory would seem to justify finite, but not countable, additivity. De Finetti gives a number of other arguments in favour of finite additivity and against countable additivity. We shall here consider one more of these.

If we adopt countable additivity, then it becomes impossible to have a uniform distribution over a countable set, such as the positive integers $\{1, 2, ..., n, ...\}$. For suppose we put $P(i) = p$ for all i. If $p > 0$, then $P(1) + P(2) + ... + P(n) + ...$ becomes infinite, whereas by the axioms of probability it should be $= 1$. If we put $P(i) = 0$ for all i, then by countable additivity $P(\{1, 2, ... , n, ...\}) = P(1) + P(2) + ... + P(n) + ... = 0$, whereas, by Axiom 1, $P(\{1, 2, ..., n, ...\}) = P(\Omega) = 1$. However, if we adopt only finite additivity, then the second half of the argument is blocked, so that it becomes possible to have a uniform distribution over the positive integers. De Finetti regards it as a counterintuitive feature of the axiom of countable additivity that it prevents us from having such uniform distributions. After all, for any finite n, however large, we can introduce a uniform distribution over the positive integers $1, 2, ..., n$ by setting $P(i) = 1/n$, $i = 1, ..., n$. However, if we postulate countable additivity over the infinite collection of positive integers $1, 2, ..., n, ...$, we can only have what he terms 'extremely unbalanced partitions' (1970:Vol. 1, 122). He explains his meaning here more fully later on when he says that countable additivity: 'forces me to choose some finite subset of them [i.e. the countable class in question, e.g. the positive integers] to which I attribute a total probability of at least 99% (leaving 1% for the remainder; and I could have said 99.999% with 0.001% remaining, or something even more extreme).' (1970:Vol. 2, 351) This argument does not perhaps go very well with the previous argument which suggests that on the subjective approach one should always limit oneself to finite collections of events and not consider probability distributions over countable sets at all.

Not all probabilists agree with De Finetti's attitude to countable additivity within the subjective theory. Adams (1964) presented a proof that countable additivity does follow from the assumptions of the subjective approach. This proof has been considerably simplified by Williamson (1999), which also discusses the philosophical problems involved. Williamson devises a betting situation in which it would seem quite reasonable to bet on a countable number of events. Suppose Ms A tells Mr B that in a sealed parcel in the next room there is the computer print-out of a positive integer, and asks him to give a betting quotient on this number being n for all n. Now of course Mr B would realise that the practicalities of technology must impose some upper bound on the value which the hidden number could take. However, this upper bound is hard to determine, and the problem is a very open-ended one. Rather than fix on a particular upper bound, it would be easier for Mr B to produce an infinite sequence of betting quotients. Actually, the infinite is often brought into applied mathematics for exactly this kind of reason.

A noteworthy feature of this example is that a uniform distribution is highly implausible. On the contrary, we would expect small numbers to be more probable than very large ones. In general, in any betting situation in which we approximate the large open-ended finite by the infinite, the unbalanced distributions described by De Finetti, far from being counterintuitive, are just what we would expect.

Williamson's other point is that, once we have introduced a betting scheme for a countably infinite number of events, it only requires one extra condition to derive the axiom of countable additivity by exactly the same Dutch book argument which De Finetti uses for finite additivity. This extra condition is that only a finite amount of money should change hands. Assuming this, let us see how the proof of Axiom 2 must be modified if we have, instead of a finite number of events $E_1, ..., E_n$, a countably infinite number $E_1, ..., E_n, ...$. Because only a finite amount of money should change hands, Ms A's gains G_i must all be finite, which means in turn that the series $q_1S_1 + ... + q_nS_n + ...$ must converge. Moreover, from Axiom 1, it follows that $q_1 + ... + q_n + ... \leq 1$. If in the proof of Axiom 2 given above, we replace the finite sums by infinite series, then, using the above results, all the series converge, and the proof goes through just as before. So, if we allow bets over a countable infinity of events (as seems eminently reasonable in the kind of situation described above), and if we specify that only a finite amount of money should change hands (which can hardly be avoided), then the axiom of countable additivity does follow rigorously from exactly the same Dutch book argument which De Finetti uses to establish finite additivity. This argument of Williamson's seems to me to show that countable additivity is completely justified within the subjective theory, and that De Finetti was wrong to deny it.

This result seems to me to strengthen rather than weaken the subjective theory. On De Finetti's approach, mathematicians who adopted the subjective theory of probability would have to use a mathematical theory somewhat different from the standard one. Many would surely regard this as an argument against becoming a subjectivist. Williamson's argument shows that such doubts are quite unnecessary, and that it is perfectly possible both to be a subjectivist and to use the standard mathematical theory. Moreover, as Williamson points out, countable additivity strengthens the subjective theory as against the logical theory. Suppose we were betting on a countably infinite sequence of events $E_1, E_2, \ldots, E_n, \ldots$, and suppose we had no reason to prefer E_i to E_j for all *i, j,* then the logical theory with its Principle of Indifference would seem to require a uniform distribution. Countable additivity forces a skew distribution on us, thus preventing a logical interpretation and introducing a subjective element. So, ironically, De Finetti's defence of a uniform distribution in this context is more of a defence of the logical view than of his own subjective approach.

Apparently objective probabilities in the subjective theory: exchangeability

So far the subjective theory has had considerable success. Starting from the analysis of probability as the degree of belief of an individual, it has shown how such degrees of belief can be measured, and how from the simple and plausible condition of coherence the standard mathematical axioms of probability can be derived. All this establishes beyond doubt that subjective probabilities are at least one of the valid interpretations of the mathematical calculus. Moreover, there are a number of situations where the subjective analysis of probability looks highly plausible. Examples would be the probability of it raining tomorrow, the probability that a particular party will win the next election or the probability of a particular horse winning a race. Such probabilities can plausibly be said to be subjective, or at least to involve a considerable subjective component. Yet there are other probabilities which do seem, at first sight at least, to be completely objective. Suppose we have a die which is shown by careful tests to be perfectly balanced mechanically, and which in a series of trials has given approximately the same frequency for each of its faces. Surely for such a die $P(5) = {}^1/_6$, and this is an objective fact, not a matter of subjective opinion. Then again consider the probability of a particular isotope of uranium disintegrating in a year. This is surely not a matter of opinion, but something which can be calculated from quantities specified in textbooks of physics. Such a probability looks every bit as objective as, for example, the mass of the isotope. How is a supporter of the subjective theory of probability to deal with cases of this sort?

Actually there are two possible approaches. First of all, it could be admitted that the examples we have cited, and others like them, are indeed objective, and consequently that there are at least two different concepts of probability which apply in different circumstances. This was the position which Ramsey (1926) adopted, and I will discuss it in Chapter 8. Second, however, it could be claimed that all probabilities are subjective, and that even apparently objective probabilities, such as the ones just described, can be explicated in terms of degree of subjective belief. This was the line adopted by De Finetti, and I will next consider his argument in detail.

De Finetti states the problem as follows:

> It would not be difficult to admit that the subjectivistic explication is the only one applicable in the case of practical predictions (sporting results, meteorological facts, political events, etc.) which are not ordinarily placed in the framework of the theory of probability, even in its broadest interpretation. On the other hand it will be more difficult to agree that this same explanation actually supplies rationale for the more scientific and profound value that is attributed to the notion of probability in certain classical domains, ...
>
> (1937:152)

Nonetheless, De Finetti does think that the subjective account of probability is adequate even in these 'classical domains', for he continues:

> Our point of view remains in all cases the same: *to show that there are rather profound psychological reasons which make the exact or approximate agreement that is observed between the opinions of different individuals very natural, but that there are no reasons, rational, positive, or metaphysical, that can give this fact any meaning beyond that of a simple agreement of subjective opinions.*
>
> (1937:152)

Let us now see how De Finetti works out this view by taking a simple example. Suppose we have a coin which is known to be biased, but for which the extent of the bias is not known. An objectivist would say that there is a true, but unknown, probability p of heads, and that we can measure p roughly by making n tosses (for large n), observing the number r of heads and setting $p \approx r/n$. The exact relation between p and r/n will depend on the particular objective theory adopted.

How then does a subjectivist like De Finetti deal with this case? The first step is to consider a sequence of tosses of the coin which we suppose gives results: $E_1, \ldots, E_n, \ldots$, where each E_i is either heads (H_i) or tails (T_i). So, in particular, H_{n+1} = Heads occurs on the n + 1th toss. Further, let e be a complete specification of the results of the first n tosses, that is a sequence n places long, at the ith place of which we have either H_i or T_i. Suppose that heads occurs r times on the first n tosses. The subjectivist's method is to calculate $P(H_{n+1} \mid e)$, and to show that under some general conditions which will be specified later $P(H_{n+1} \mid e)$ tends to r/n for large n. This shows that whatever value is assigned to the prior probability $P(H_{n+1})$, the posterior probability $P(H_{n+1} \mid e)$ will tend to the observed frequency for large n. Thus, different individuals who may hold widely differing opinions initially will, if they change their probabilities by Bayesian conditionalisation, come to agree on their posterior probabilities. The objectivist wrongly interprets this as showing that there is an objective probability, but, according to De Finetti, 'objective probability' is a metaphysical concept devoid of meaning. All that is happening is that, in the light of evidence, different individuals are coming to agree on their subjective probabilities. Such is the argument. Let us now give, in our simple case, the mathematical proof which underpins it.

Suppose that $P(E_i) \neq 0$ for all *i*, so that also $P(e) \neq 0$. We then have by Axiom 3

$$P(H_{n+1} \mid e) = \frac{P(H_{n+1} \,\&\, e)}{P(e)} \tag{4.3}$$

To proceed further we introduce the condition of *exchangeability.* Suppose Mr B is making an a priori bet that a particular *n*-tuple of results (E_{i1} E_{i2} ... E_{in} say) occurs. Suppose further that heads occurs *r* times in this *n*-tuple. Mr B's betting quotients are said to be *exchangeable* if he assigns the same betting quotient to any other particular *n*-tuple of results in which heads occurs *r* times, where both *n* and *r* can be chosen to have any finite integral non-negative values with $r \leq n$. Let us write his prior probability (or betting quotient) that there will be *r* heads in *n* tosses as $\omega_r^{(n)}$. There are ${}^nC_r$ different ways in which *r* heads can occur in *n* tosses, where, as usual, ${}^nC_r = n!/(n - r)!\, r! = n(n - 1) \ldots (n - r + 1)/r(r - 1) \ldots 1$. Each of the corresponding *n*-tuples must, by exchangeability, be assigned the same probability, which is therefore $\omega_r^{(n)}/{}^nC_r$. Thus

$$P(E_{i1}E_{i2}\ldots E_{in}) = \frac{\omega_r^n}{{}^nC_r} \tag{4.4}$$

Now e, by definition, is just a particular *n*-tuple of results in which heads occurs *r* times. Thus, by exchangeability,

$$P(e) = P(E_{i1}E_{i2}\ldots E_{in}) = \frac{\omega_r^n}{{}^nC_r} \tag{4.5}$$

Now H_{n+1} & e is an $(n + 1)$-tuple of results in which heads occurs $r + 1$ times. Thus, by the same argument,

$$P(H_{n+1} \,\&\, e) = \frac{\omega_{r+1}^{(n+1)}}{{}^{n+1}C_{r+1}} \tag{4.6}$$

And so, substituting in Equation 4.3, we get

$$\begin{aligned} P(H_{n+1} \mid e) &= \frac{{}^nC_r}{{}^{n+1}C_{r+1}} \frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} \\ &= \frac{n!}{(n-r)!\,r!} \frac{(r+1)!\,(n-r)!}{(n+1)!} \frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} \\ &= \frac{r+1}{n+1} \frac{\omega_{r+1}^{(n+1)}}{\omega_r^{(n)}} \end{aligned} \tag{4.7}$$

Equation 4.7 gives us the result we want. Provided only $\omega_{r+1}^{(n+1)}/\omega_r^{(n)} \to 1$ as $n \to \infty$ (a very plausible requirement), we may choose our prior probabilities $?_r^{(n)}$ in any way we please, and still get that as $n \to \infty$, $P(H_{n+1} \mid e) \to r/n$ (the observed frequency), as required.

To sum up then: according to the objectivist, there is a real objective probability p of heads, and the observed frequency r/n gives an increasingly better estimate of p as $n \to \infty$.

According to the subjectivist, the 'real objective probability p' is a metaphysical delusion. Different people may, subject only to coherence, have different prior probabilities $P(H_{n+1})$. However, coherence + exchangeability + one other plausible assumption ($\omega_{r+1}^{(n+1)}/\omega_r^{(n)} \to 1$, as $n \to \infty$) ensure that $P(H_{n+1} \mid e) \to r/n$ as $n \to \infty$. Thus, as the evidence piles up, the people who disagree a priori will come to agree a posteriori. This 'exact or approximate agreement between the opinions of different individuals for rather profound psychological reasons' is what gives rise to the illusion of objective probabilities.

In n tosses, we can have either 0, 1, 2, ..., or n heads. So, by coherence,

$$\omega_0^{(n)} + \omega_1^{(n)} + \omega_2^{(n)} + \ldots + \omega_r^{(n)} + \ldots + \omega_n^{(n)} = 1 \tag{4.8}$$

In the subjective theory, we can choose the $\omega_r^{(n)}$ (the prior probabilities) in any way we choose subject only to Equation 4.8. However, we can also, though this is not compulsory, make the 'Principle of Indifference' choice of making them all equal so that

$$\omega_o^{(n)} = \omega_1^{(n)} = \omega_2^{(n)} = \ldots = \omega_r^{(n)} = \ldots = \omega_n^{(n)} = 1/(n+1) \tag{4.9}$$

Substituting this in Equation 4.7, we get

$$P(H_{n+1} \mid e) = \frac{r+1}{n+2} \tag{4.10}$$

This is a classical result – Laplace's Rule of Succession.

The Rule of Succession has been used to try to solve Hume's problem of induction. Suppose, having read Hume, we are worried about whether the Sun will rise tomorrow. Now recorded history goes back at least 5,000 years, and the Sun has been observed (in the appropriate latitudes) to rise every single morning during all that time. At least, if the Sun had failed to rise one morning, it is a reasonable presumption that this fact would have been recorded. So our evidence is that the Sun has risen each morning for 1,826,250 days. To calculate the probability of its rising tomorrow, we use Equation 4.10 with $r = n = 1{,}826{,}250$. This gives the probability of the Sun's rising tomorrow as approximately 0.9999994. If this reasoning is correct, then we should no longer be troubled by Humean doubts, but should be able to look forward with very great confidence to the Sun rising tomorrow!

But not everyone is convinced by the argument, and the Rule of Succession has been subjected to quite a number of harsh criticisms. I will here describe one based on an example due to Popper.[7] Suppose the inhabitants of London wake up one summer morning to find that although according to their clocks it should be day, it is in fact still night outside. They switch on their radios and televisions and learn that something quite extraordinary has happened. The Earth appears to have stopped rotating. It is still night in London, while on the opposite side of the globe, the Sun is staying fixed at one position in the sky. Of course this quite contradicts all the known laws of physics. Moreover, apart from the strange change in the apparent movements of the Sun, everything else seems to be continuing just as before, a situation which again contradicts all the known laws of physics. Scientists the world over confess that they are baffled and cannot understand what is happening. Copies of the philosophical works of Hume are selling well.

Given this bizarre, but at least imaginable, situation, what would be the probability of the Sun's rising again as usual the next morning? It is easy to calculate according to the Rule of Succession. In Equation 4.10 above, we now have $r = n - 1$, and $n = 1,826,251$. So the probability of the Sun's rising the next day is 0.9999989. In other words, if we stick to the Rule of Succession, the quite extraordinary events just described would reduce the probability of the Sun's rising the next day by 0.0000005, i.e. 5×10^{-7}. Obviously this is quite wrong. There would be such a state of confusion that no one would have the least idea of whether the Sun would rise the next day or not. Certainly no one would assign a probability of 0.9999989 to its doing so. This example shows that, although the Rule of Succession may give reasonable answers in some cases, it gives absurd answers in others and so cannot be considered valid in general. On the other hand, it is not clear what exactly is wrong with the rather convincing chain of reasoning which was presented above and which led to the Rule of Succession. Rather than pursue this problem immediately, I will first present a general criticism of De Finetti's analysis of apparent objectivity in terms of exchangeability. This criticism casts light on why the Rule of Succession fails so dramatically in some cases, as I will then show.

To explain my criticism of De Finetti's exchangeability argument, I will begin by quoting an important passage in which he describes some general features of the argument. It is precisely these features which I will then criticise. The passage runs as follows:

> Whatever be the influence of observation on predictions of the future, it never implies and never signifies that we *correct* the primitive evaluation of the probability $\mathbf{P}(E_{n+1})$ after it has been *disproved* by experience and substitute for it another $\mathbf{P}^*(E_{n+1})$ which *conforms* to that experience and is therefore probably *closer to the real probability*; on the contrary, it manifests itself solely in the sense that when experience teaches us the result A on the first n trials, our judgment will be expressed by the probability $\mathbf{P}(E_{n+1})$ no longer, but by the probability $\mathbf{P}(E_{n+1} \mid A)$, i.e. that which our initial opinion would already attribute to the event E_{n+1} considered as conditioned on the outcome A. Nothing of this initial opinion is repudiated or corrected; it is not the

> function **P** which has been modified (replaced by another P*), but rather the argument E_{n+1} which has been replaced by $E_{n+1} | A$, and this is just to remain faithful to our original opinion (as manifested in the choice of the function P) and coherent in our judgment that our predictions vary when a change takes place in the known circumstances.
>
> In the same way, someone who has the number 2374 in a lottery with 10,000 tickets will attribute at first a probability of 1/10,000 to winning the first prize, but will evaluate the probability successively as 1/1000, 1/100, 1/10, 0, when he witnesses the extraction of the successive chips which give, for example, the number 2379. At each instant his judgment is perfectly coherent, and he has no reason to say at each drawing that the preceding evaluation of probability was not right (at the time when it was made).
>
> (De Finetti 1937:146–7)

This passage puts very clearly the difference between De Finetti's position and that of an objectivist – particularly an objectivist with Popperian tendencies. For such an objectivist, any evaluation ***P*** of a probability function is just a conjecture as to the values of the real objective probabilities, and, like any conjecture it should be severely tested. If these tests show that it is inadequate in anyway, it should be replaced by a new conjecture ***P****** which fits the facts better. In De Finetti's scheme, we do not try to test or refute our prior probabilities **P**(E_{n+1}), we simply change them into posterior probabilities P(En + 1 | A) by Bayesian conditionalisation. Different people may start with different prior probabilities, but, as the evidence mounts up, their posterior probabilities will tend in many circumstances to converge producing the illusion of the existence of an objective probability.

My argument against De Finetti can be stated in general terms as follows. The prior probability function ***P*** will in all cases be based on general assumptions about the nature of the situation under study. Now if these assumptions are broadly correct, then De Finetti's scheme of modifying ***P*** by Bayesian conditionalisation will yield reasonable results. If, however, the initial assumptions are seriously wrong in some respects, then not only will the prior probability function be inappropriate, but all the conditional probabilities generated from it in the light of evidence will also be inappropriate. To obtain reasonable probabilities in such circumstances, it will be necessary to change ***P*** in a much more drastic fashion than De Finetti allows, and, in effect, introduce a new probability function ***P*****. This line of thought could be summarised as follows. De Finetti's scheme of allowing changes only by Bayesian conditionalisation is too conservative. Sometimes, in order to make progress, much more drastic changes in ***P*** are needed than those which he allows. I will give an example of such a situation in a moment. However, to explain the general character of this example, it will be desirable to examine the relation between the concepts of independence and exchangeability. As this involves some technicalities I will discuss the matter in the next section. I will then give an informal summary of the main points of this section before giving my example of a situation in which De Finetti's method of changing prior probabilities only by Bayesian conditionalisation proves to be inadequate.

The relation between independence and exchangeability*

In a certain sense the concept of exchangeability is the equivalent within the subjective theory of the objectivist's notion of independence. This does not mean that the concept of independence does not apply in the subjective theory. Two events E, F are defined to be independent, if P(E & F) = P(E) P(F). This definition can of course be applied when the probabilities involved are given a subjective meaning. The trouble is that while in objective approaches the assumption of independence is a very important one which applies in many cases, independence in the subjective sense turns out to be an assumption which can rarely, if ever, be made. If we make the mathematical assumption of independence, giving the probabilities an epistemological meaning, this turns out to give a case in which no learning from experience can occur. We can see this in the context of the subjective theory by exploring what happens if we change the assumption of exchangeability to that of independence. This amounts to assuming that

$$\mathrm{P}(\mathrm{E}_{i1} \;\&\; \mathrm{E}_{i2} \;\&\; \ldots \;\&\; \mathrm{E}_{in}) = \mathrm{P}(\mathrm{E}_{i1})\, \mathrm{P}(\mathrm{E}_{i2}) \ldots \mathrm{P}(\mathrm{E}_{in})$$

It follows in particular that $\mathrm{P}(\mathrm{H}_{n+1} \;\&\; \mathrm{e}) = \mathrm{P}(\mathrm{H}_{n+1})\,\mathrm{P}(\mathrm{e})$. Substituting this into Equation 4.3 above, we get

$$\mathrm{P}(\mathrm{H}_{n+1} \mid \mathrm{e}) = \mathrm{P}(\mathrm{H}_{n+1})$$

So within the Bayesian framework no learning from experience can occur. De Finetti must have realised this very early on in his development of the subjective theory for he writes:

> If the outcome of the preceding trials can modify my opinion, it is for me *dependent* and *not* independent.... If I admit the possibility of modifying my probability judgment in response to observation of frequencies; it means that – by definition – my judgment of the probability of one trial is not independent of the outcomes of the others ....
>
> (1931a:212)

In general, an individual such as our Mr B will want to modify his probability judgements in response to observation of frequencies, and so it follows that the assumption of independence will rarely, if ever, be made within the subjective theory. At first sight this may seem rather a severe blow to the subjective approach, since objectivists frequently and successfully make assumptions of independence. This was no doubt one factor which stimulated De Finetti to invent his new concept of exchangeability. Roughly speaking where an objectivist assumes independence, a subjectivist will assume exchangeability. De Finetti proved a general theorem showing how the two concepts are linked, I will next state his result.

Let us first define exchangeability for a sequence of random variables (or random quantities as De Finetti prefers to call them) $X_1, \ldots, X_n, \ldots$. These are exchangeable if, for any fixed n, X_{i1}, X_{i2}, ..., X_{in} have the same joint distribution no matter how $i1$, ..., in are chosen. Now let Y_n be the average of any n of the random quantities X_i, i.e. $Y_n = (1/n)(X_{i1} + X_{i2} + ... + X_{in})$, since we are dealing with exchangeable random quantities it does not matter which $i1$, $i2$,..., in are chosen. De Finetti first shows (1937: 126) that the distribution $\Phi_n(\xi) = P(Y_n \leq \xi)$ tends to a limit $\Phi(\xi)$ as $n \to \infty$, except perhaps for points of discontinuity. He goes on to say:

> Indeed, let $\mathbf{P}_\xi$(E) be the probability attributed to the generic event E when the events E_1, E_2, ..., E_n, ... are considered independent and equally probable with probability ?; the probability **P**(E) of the same generic event, the E_i being exchangeable events with the limiting distribution $\Phi(\xi)$, is
>
> $$\mathbf{P}(E) = \int_0^1 \mathbf{P}_\xi(E) d\Phi(\xi)$$
>
> This fact can be expressed by saying that the probability distributions **P** corresponding to the case of exchangeable events are linear combinations of the distributions $\mathbf{P}_\xi$ corresponding to the case of independent equiprobable events, the weights in the linear combination being expressed by $\Phi(\xi)$.
>
> (De Finetti 1937:128–9)

This general result can be illustrated by taking a couple of special cases. Suppose that we are dealing with a coin-tossing example and the generic event E is that heads occurs r times in n tosses. Then

$$\mathbf{P}_\xi(E) = {}^nC_r \xi^r (1 - \xi)^{n - r}$$

So

$$\mathbf{P}(E) = \omega_r^{(n)} = {}^nC_r \int_0^1 \xi^r (1 - \xi)^{n-r} d\mathbf{\Phi}(\xi)$$

If, in particular, F(?) is the uniform distribution, we have

$$\omega_r^{(n)} = {}^nC_r \int_0^1 \xi^r (1 - \xi)^{n-r} d\mathbf{\Phi}(\xi)$$

$= {}^{n}C_{r}\, B(r + 1, n - r + 1)$, where B is the beta function

$= 1/(n + 1)$ (cf. Equation 4.9 above)

Comparing these results with our earlier calculations involving exchangeability, we can see how exchangeability and independence are related.

De Finetti interprets these mathematical results as showing that we can *eliminate* the notions of objective probability and independence (which in his view are metaphysical in character) in favour of those of subjective probability and exchangeability. Alternatively, we could speak of his results as a *reduction* of objective probability and independence to subjective probability and exchangeability. The idea is that when an objectivist assumes independence, and formulates corresponding mathematical equations, a subjectivist can simply reinterpret these equations as being about subjective probabilities and exchangeability. This interpretation eliminates the objectivist's metaphysical notions and gives the real empirical meaning of the equations. I will call this argument *De Finetti's exchangeability reduction* and will criticise it in the next section.

Criticism of De Finetti's exchangeability reduction

In the previous section, it has been shown that exchangeability is in a sense the subjective equivalent of objective independence. De Finetti takes this to mean that we can eliminate the objectivist's notion of independence in favour of exchangeability. From the objectivist's point of view, however, the relation can be read, so to speak, in the opposite direction as showing that we can only apply exchangeability *when the situation is objectively one of independence.* However, not all sequences of events are independent. On the contrary, there are many situations in which the outcome of a particular event is very strongly dependent on the outcomes of previous events. In such situations we would expect that the use of exchangeability, and the calculations with it explained above, would give completely erroneous results. This is indeed the case, as I will illustrate in a moment by means of an example. My conclusion is that far from our being able to reduce the notion of objective independence to that of exchangeability, the concept of exchangeability is actually parasitic on that of objective independence and so redundant. In order to use exchangeability in a way which does not lead to erroneous and misleading results, we have first to be sure that the situation is objectively one of independence. We can only acquire such a conviction by conjecturing that the situation is one of independence and testing this assumption rigorously. If our conjecture passes these tests, then we can use the exchangeability calculation without going far wrong, but there is no need to do so, since we handle the problem in the standard way, using independence and objective probabilities. In this case then, exchangeability is unnecessary. If, on the other hand, our tests show that the situation is not one of independence, then the use of exchangeability will give misleading results and should be avoided. In neither case therefore is there any reason for using exchangeability.

To illustrate this argument, it would be possible to use any sequence of events which are dependent rather than independent. I have chosen one very simple and at the same time striking example of dependence. This is the game of red or blue.[8] At each go of the game there is a number s which is determined by the previous results. A fair coin is tossed. If the result is heads, we change s to $s' = s + 1$, and if the result is tails, we change s to $s' = s - 1$. If $s' \geq 0$, the result of the go is said to be blue, whereas if $s' < 0$, the result of the go is said to be red. So, although the game is based on coin tossing, the results are a sequence of red and blue instead of a sequence of heads and tails. Moreover, although the sequence of heads and tails is independent, the sequence of red and blue is highly dependent. We would expect much longer runs which are all blue than runs in coin tossing which are all heads. If we start the game with $s = 0$, then there is a slight bias in favour of blue, which is the initial position. However, it is easy to eliminate this by deciding the initial value of s by a coin toss. If the toss gives heads we set the initial value of s at 0, and if the toss gives tails we set it at -1. This makes red and blue exactly symmetrical, so that the limiting frequency of blue must equal that of red and be $^1/_2$. It is therefore surprising that over even an enormously large number of repetitions of the game, there is high probability of one of the colours appearing much more often than the other. Feller (1950:82–3) gives a number of examples of these curious features of the game. Suppose for example that the game is played once a second for a year, i.e. repeated 31,536,000 times. There is a probability of 70 per cent that the more frequent colour will appear for a total of 265.35 days, or about 73 per cent of the time, whereas the less frequent colour will appear for only 99.65 days, or about 27 per cent of the time.

Let us next suppose that two probabilists – an objectivist (Ms A) and a subjectivist (Mr B) – are asked to analyse a sequence of events, each member of which can have one of two values. Unknown to them, this sequence is in fact generated by the game of red or blue. Possibly the sequence might be produced by a man-made device which flashes either 0 (corresponding to red) or 1 (corresponding to blue) on to a screen at regular intervals. However, it is not impossible that the sequence might be one occurring in the world of nature. Consider for example a sequence of days, each of which is classified as 'rainy' if some rain falls, or dry otherwise. In a study of rainfall at Tel Aviv during the rainy season of December, January and February, it was found that the sequence of days could be modelled successfully as a sequence of dependent events. The particular kind of dependence used was what is known as a *Markov chain,* that is to say the probability of a day being rainy was postulated to depend on the weather of the previous day, but not on the weather of days further back in the sequence. In fact, the probabilities found empirically were probability of a dry day given that the previous day was dry = 0.75, and probability of a rainy day given that the previous day was rainy = 0.66. (For further details see Cox and Miller 1965:78–9.) It is clear that this kind of dependence will give longer runs of either rainy or dry days than would be expected on the assumption of independence. It is thus not impossible that the sequence of rainy and dry days at some place and season might be represented quite well by the game of red or blue.

Let us return to our two probabilists and consider first the objectivist (Ms A). Knowing that the sequence has a random character, she will begin by making the simplest and most familiar conjecture that the events are independent. However, being a good Popperian, she will test this conjecture rigorously with a series of statistical tests for independence. It will not be long before she has rejected her initial conjecture, and she will then start exploring other hypotheses involving various kinds of dependence among the events. If she is a talented scientist, she may soon hit on the red or blue mechanism and be able to confirm that it is correct by another series of statistical tests.

Let us now consider the subjectivist Mr B. Corresponding to Ms A's initial conjecture of independence, he will naturally begin with an assumption of exchangeability. Let us also assume that he gives a uniform distribution a priori to the $\omega_r^{(n)}$ (see Equation 4.9 above) so that Laplace's Rule of Succession holds (Equation 4.10). This is just for convenience of calculation. The counterintuitive results would appear for any other coherent choice of the $\omega_r^{(n)}$. Suppose that we have a run of 700 blues followed by two reds. Mr B would calculate the probability of getting blue on the next go using Equation 4.10 with $n = 702$ and $r = 700$. This gives the probability of blue as ${}^{701}/_{704} = 0.996$ to three significant figures. Knowing the mechanism of the game, we can calculate the true probability of blue on the next go, which is very different. Go 700 gave blue, and go 701 gave red. This is only possible if s on go 700 was 0, the result of the toss was tails and s became -1 on go 701. The next toss must also have yielded tails or there would have been blue again on go 702. Thus s at the start of go 703 must be -2, and this implies that the probability of blue on that go is zero. Then again let us consider one of Feller's massive sessions of 31,536,000 goes. Suppose the result is that the most frequently occurring colour appears 73 per cent of the time (as pointed out above there is a probability of 70 per cent of this result, which is thus not an unlikely outcome). Mr B will naturally be estimating the probability of this colour at about 0.73 and so much higher than that of the other colour. Yet in the real underlying game, the two colours are exactly symmetrical.

We see that Mr B's calculations using exchangeability will give results at complete variance with the true situation. Moreover, he would probably soon notice that there were too many long runs of one colour or the other for his assumption of exchangeability to be plausible. He might therefore think it desirable to change his assumption of exchangeability into some other assumption. Unfortunately, however, he would not be allowed to do so according to De Finetti, for, to quote again a section of the key passage given above:

> ... when experience teaches us the result A on the first n trials, our judgment will be expressed by the probability $\mathbf{P}(E_{n+1})$ no longer, but by the probability $\mathbf{P}(E_{n+1} | A)$, i.e. that which our initial opinion would already attribute to the event E_{n+1} considered as conditioned on the outcome A. Nothing of this initial opinion is repudiated or corrected; it is not the function $\mathbf{P}$ which has

> been modified (replaced by another **P***), but rather the argument E_{n+1} which has been replaced by $E_{n+1} \mid A$, and this is just to remain faithful to our original opinion (as manifested in the choice of the function **P**) ...
>
> (1937:146)

Yet if we assume exchangeability a priori when the sequence of events is in reality dependent, no amount of modifying our prior probabilities $P(E_{n+1})$ to posterior probabilities $P(E_{n+1} \mid A)$ by Bayesian conditionalisation will produce probabilities which accord with the real situation. De Finetti's exchangeability analysis only looked plausible in the first place because it was applied to coin tossing, and we know from long experience that tosses of a coin can validly be considered to be objectively independent. Unless we know that the events are objectively independent, we have no guarantee that the use of exchangeability will lead to reasonable results.

This point explains why the Rule of Succession leads to such erroneous results in the case in which the Sun mysteriously fails to rise one morning. Of course our background knowledge tells us that successive risings of the Sun are not independent events, but are highly dependent. This explanation of the situation can be reinforced by considering a case in some respects like the example of the Sun rising, but in which we do know that the events are independent. In such a case, as we shall see, the Rule of Succession gives perfectly reasonable results.

Suppose we have a large number of balls in a container. The container is thoroughly shaken, a ball is drawn, its colour is noted and it is then replaced. We can suppose that, as part of our background knowledge, we have a detailed acquaintance with all the mechanisms involved so that we can be sure that the drawings are independent. We do not, however, know the number of balls in the container or their colour. In fact, there are 1,000,000 balls of which 999,999 are yellow (corresponding to the Sun rising), and one is black (corresponding to its failing to rise). Suppose a yellow ball is drawn 737,856 times, and then a black ball is drawn. The Rule of Succession gives 737,856/737,858 = 0.9999972 to seven significant figures for the probability of drawing a yellow ball on the next occasion. This is actually not unreasonable in the circumstances. The results so far indicate that there must be an overwhelming preponderance of yellow balls in the container. So that, even if there are a few black balls, we are still much more likely to get a yellow ball on the next draw, provided the container is shaken very thoroughly (independence assumption). The Rule of Succession gives a reasonable result in this case of drawing balls from a container, but an absurd result in the case of the Sun failing to rise. This is because we know that independence applies in the case of drawing the balls, and that it doesn't apply in the case of the Sun either rising or failing to rise. This reinforces our conclusion that we can only apply exchangeability if we are sure on the basis of our background knowledge that the events concerned are objectively independent.

This concludes my criticism. Let us now see how a supporter of De Finetti might try to answer it. De Finetti himself does say one or two things which are relevant to the problem. Having shown that exchangeable events are the subjective equivalent of the objectivist's independent and equiprobable events, he observes that one could introduce subjective equivalents of various forms of dependent events, and, in particular, of Markov chains. As he says:

> One could in the first place consider the case of classes of events which can be grouped into Markov "chains" of order 1,2, ..., m, ..., in the same way in which classes of exchangeable events can be related to classes of equiprobable and independent events.
>
> (De Finetti 1937: Footnote 4, 146)

We could call such classes of events *Markov exchangeable.* De Finetti argues that they would constitute a complication and extension of his theory without causing any fundamental problem:

> One cannot exclude completely *a priori* the influence of the order of events.... There would then be a number of degrees of freedom and much more complication, but nothing would be changed in the setting up and the conception of the problem ..., before we restricted our demonstration to the case of exchangeable events ...
>
> (1937:145)

Perhaps De Finetti has in mind something like the following. Instead of just assuming exchangeability, we consider not just exchangeability but various forms of Markov exchangeability. To each of these possibilities we give a prior probability. No doubt exchangeability will have the highest prior probability. If the case is a standard one, like the biased coin, this high prior probability will be reinforced, and the result will come out moreover less like that obtained by just assuming exchangeability. If, however, the case is an unusual one, then the posterior probability of exchangeability will gradually decline, and that of one of the other possibilities will increase until it becomes much more probable than exchangeability. Does a scheme of this sort resolve the problems which have been raised? I will now argue that it does not.

The main problem with the approach just sketched is that it is unworkably complicated, and moreover these complications are quite unnecessary since they can be eliminated completely on the objective approach. I will deal with these points in turn. What leads to so much complication is that on this approach *it is necessary to consider all the possibilities which might arise at the very beginning of the investigation.* In order to set up his prior probabilities, Mr B has to consider every possible kind of dependence which might arise in the sequence of events, and assign each a prior probability. Now there is a very large number of different forms of dependence. De Finetti mentions Markov chains of different orders, but there are non-Markovian forms of dependence as well. Even if Mr B listed all the forms of dependence which have been so far explicitly defined and studied by mathematicians, he could still miss the one which applies to the sequence of events he is considering because this might be of a hitherto unstudied form. Yet for Mr B to list and assign prior probabilities to all forms of dependence known at present would be a task of such complexity so as to exceed most human powers. It is a testimony to the difficulty of this task that no one has, to my knowledge, carried it out in detail. Moreover, and this is my second point, all this complication is eliminated completely by adopting the objective approach. Our objectivist Ms A, when considering a sequence of events of a hitherto unstudied type, need only consider a single possibility to begin with. She could start with the conjecture that the events are independent with constant probabilities for the various outcomes. She does not need to bother a priori with other hypotheses of dependence, variable probabilities, or whatever, because, being a good Popperian, *she will subject her initial conjecture to a series of rigorous statistical tests.* Perhaps these tests will corroborate her initial conjecture in which case an elaborate a priori consideration of other possibilities would have been a waste of time and trouble. Perhaps, however, the test will refute her conjecture, in which case she will, at that stage and in the light of the results obtained, attempt to devise some new hypothesis. By approaching the problem in this step-by-step fashion, it is rendered tractable, whereas the Bayesian attempt to consider all possibilities a priori is quite unworkable.

Let us now consider another way in which the criticism we have made might be answered. A subjectivist might argue that De Finetti's requirement that prior probabilities should be changed only by Bayesian conditionalisation, i.e. from $\mathbf{P}(E_{n+1})$ to $\mathbf{P}(E_{n+1} \mid A)$ is too strong. Maybe prior probabilities should generally be altered in this fashion, but perhaps if exceptional results appear, as in the game of red or blue, prior probabilities could be altered in some quite different fashion to take account of the new circumstances. This solution of the difficulty certainly appeals to common sense, and would, I am sure, be adopted in practice. Unfortunately, however, it destroys the basis of De Finetti's exchangeability reduction, and even of Bayesianism in general. The exchangeability reduction works by arguing that whatever prior probabilities a set of different people adopt, their posterior probabilities will converge towards the same value. However, this argument is only valid on the assumption that all members of the set are changing their prior probabilities to posterior probabilities by Bayesian conditionalisation. If they are allowed at any time to change their priors in some quite different fashion (as on the present suggestion), there is no guarantee that their posterior probabilities will become at all similar. After 500 events, Mr B might suddenly decide to change to some form of Markov exchangeability, while Ms C continues to use exchangeability. After 700 events their posterior probabilities could be completely different. Moreover, it is one of the most attractive features of Bayesianism that it offers a simple mathematical formula for the way in which a rational person should change his or her beliefs in the light of evidence. If we now say: 'well, sometimes rational people should use this mathematical formula to change their beliefs, but, of course, it is quite open to them whenever they feel like it to change their beliefs in a completely different way', then surely we have lost that very feature which made Bayesianism an appealing theory.

I conclude that De Finetti's exchangeability reduction does not work, and it will be obvious that my arguments against this reduction can be used against Bayesianism in general. I will consider this matter briefly in the next section.

Some objections to Bayesianism

Most Bayesian statisticians use Bayesianism in something like the following form. They suppose that, in a given problem, there is a set of possible hypotheses to be considered. This set can be written $\{H_\theta\}$ where $\theta \varepsilon$ I, for some set I, usually an interval of the real line. The parameter θ is given a prior distribution $\mu(\theta)$ say, and this is changed to a posterior distribution $\mu(\theta \mid e)$. These distributions are in effect over the set of hypotheses under consideration. So let us set $P(H_\theta) = \mu(\theta)$ and $P(H_\theta \mid e) = \mu(\theta \mid e)$.

We can test this approach using the following simple 'black box' model. Mr B is confronted with a black box which flashes a figure (either 0 or 1) on to a screen at regular intervals $t = 0, 1, 2, ..., n, ...$ Let the sequence of figures be $x_0, x_1, x_2, ..., x_n, ...$ It is generated by some process unknown to Mr B. Mr B has to assign probabilities of the form $P(x_n \mid x_1, x_2, ..., x_{n-1})$ when he knows the value of $x_0, x_1, x_2, ..., x_{n-1}$ but not that of x_n. These probabilities are taken as his betting quotients in the usual gambling game played with Ms A on the value of the *n*th figure. Mr B tackles this problem by using the standard approach of a Bayesian statistician described in the first paragraph of this section. If e states the observed values of $x_0, x_1, x_2, ..., x_{n-1}$, he uses $P(H_\theta \mid e)$ to calculate $P(x_n \mid e)$.

In this framework, we can restate the objection, based on the game of red or blue, and given previously (p. 79). Suppose Mr B chooses H_θ = the sequence is independent with $\text{Prob}(1) = \theta$, $0 \leq \theta \leq 1$. Suppose further that the sequence is in reality generated by the game of red or blue with red = 0, blue = 1. Arguing as in the previous section, we can show that Mr B's systematic use of Bayesian conditionalisation as his means of learning will produce a sequence of probabilities at complete variance with reality. Bayesian conditionalisation will not therefore be a very effective learning strategy.

The obvious reply which a Bayesian might make to this argument is that Mr B has considered too narrow a class of hypotheses and a broader class should have been introduced. Albert has, however, shown that there is a serious difficulty with this reply.[9] Albert asks us to suppose that the 0s and 1s flashing on the screen of the black box are generated by what he calls a *Chaotic Clock.* This device is illustrated in Figure 4.1. There is one pointer that can point to all real numbers in the interval I = [0, 1], where the vertically upward position is zero and the vertically downward position is $^1/_2$. Initially, the pointer deviates by an angle $\omega = 2\theta\pi$ from the vertically upward position, thus pointing at the real number θ. At $t = 1, 2, ..., n, ...$, the pointer moves by doubling the angle ω.

In terms of the chaotic clock, Mr B can form hypotheses as to how the sequence of 0s and 1s is generated. H_θ might be that θ is the initial position of the pointer and that if the pointer comes to rest in the left hand side of the dial, the screen of the black box shows 0, while otherwise it shows 1. For technical reasons, Albert (1999) considers a slight modification of this chaotic clock set of hypotheses, and Suppose Mr B adopts any learning strategy whatever, i.e. he chooses his

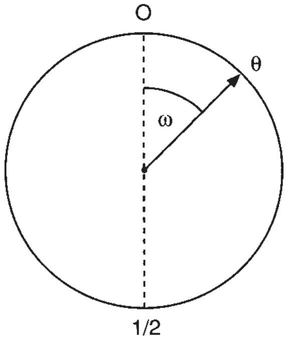


Figure 4.1 A chaotic clock

sequence of $P(x_n \mid e)$ in any arbitrary way. There then exists a prior probability distribution μ over the set of modified chaotic clock hypotheses such that Mr B's probabilities are produced by Bayesian conditioning of μ.

Albert's result is very striking indeed. His chaotic clock hypotheses are by no means absurd. After all, chaos theory is used in both physics and in economics. Indeed, hypotheses involving chaos are quite plausible as a means of explaining, for example, stock market fluctuations. If Mr B were really faced with a bizarre sequence of 0s and 1s, why should he not consider a hypothesis based on chaos theory? His imaginary situation is not so very different from the real situation of traders in financial markets who sit glued to their computer screens and make bets on what will appear shortly. Yet if Mr B is allowed to consider the chaotic clock set of hypotheses, then any learning strategy he adopts becomes a Bayesian strategy for a suitable choice of priors. In effect, Bayesianism has become empty.

It follows that a Bayesian of the type we are considering in this section (Mr B say) is caught on the horns of a dilemma. Mr B may adopt a rather limited set of hypotheses to perform his Bayesian conditionalisation, but then, as the example of the game of red or blue shows, if his set excludes the true hypothesis his Bayesian learning strategy may never bring him close to grasping what the real situation is. This is the first, or 'red or blue', horn of the dilemma. If Mr B responds by saying he is prepared to consider a wide and comprehensive set of hypotheses, these will surely include hypotheses from chaos theory and thus anything he does will become Bayesian, making the whole approach empty. This is the second, or 'chaotic clock', horn of the dilemma.

These difficulties with Bayesianism and, more specifically, with De Finetti's exchangeability reduction do indicate that there may be a need for objective probabilities and a methodology for statistics based on testing. This is therefore a good point at which to begin considering the principal objective theories of probability which will be dealt with in the next three chapters. I will, however, conclude the present chapter by considering in the last section the historical background to De Finetti's introduction of the subjective theory.

De Finetti's route to subjective probability

Earlier (pp. 52–3) I showed how Ramsey arrived at the subjective theory of probability through a criticism of Keynes's logical theory. This was not the way that De Finetti came to the theory, however, since, as I pointed out earlier, he only studied Keynes's views on probability carefully after he had already formulated the subjective theory. But what then was De Finetti's route to subjective probability?

De Finetti (1995) gives some reminiscences about when he first concluded that probability was subjective. As far as he could remember, the adoption of this philosophical position occurred very early in his intellectual career, and in fact:

> When I was a student, probably two years before graduating, while I was studying a book of Czuber's, *Wahrscheinlichkeitsrechnung* ... In that book there was a brief account of the various conceptions of probability, presented very sketchily in the first few paragraphs. Now I don't remember well the contents of the book either in general or regarding the various conceptions of probability. It seems to me that he mentioned De Morgan as representative of the subjective point of view.... Comparing the various positions it seemed to me that all the other definitions were meaningless. In particular the definition which is based on the so-called "equally probable cases" seemed to me unacceptable.
>
> (De Finetti 1995:111)

Czuber's book on probability was published in 1903, with a second enlarged and revised edition appearing in 1908–10. It was an important work in the early decades of the twentieth century and is referred to extensively by Keynes. It is worth noting that Keynes states that Czuber gives one of the best accounts of the paradoxes of geometrical probability (Keynes 1921:47), but that nonetheless Czuber thought that some form of the Principle of Non-sufficient Reason was indispensable.

In De Finetti's (1931a) first systematic account of the philosophy of probability, there are, however, no references to either Czuber or De Morgan. Instead, he cites mainly the writings of the French school of probabilists: Bertrand, Borel, Lévy and Poincaré. These writers were of course steeped in the Laplacean tradition, and their writings (particularly those of Bertrand and Borel) contained detailed discussions of the paradoxes of the Principle of Indifference. Thus, although De Finetti's reading must have been considerably different from Ramsey's, he was faced with the same problem situation – namely the difficulties for the traditional Laplacean kind of Bayesianism created by paradoxes of the Principle of Indifference. These paradoxes arose because of the perceived need to generate a single correct probability by some kind of logical process. They are thus resolved by the subjective move which allows different people to have different prior probabilities without this creating a contradiction.

However, De Finetti does not focus narrowly on the problems generated by the Principle of Indifference, but he rejects the whole Laplacean outlook, both

Laplace's determinism and his acceptance of the enlightenment value of rationality. Regarding determinism, De Finetti says:

> Certainly, we cannot accept determinism; we cannot accept the "*existence*", in that famous alleged realm of darkness and mystery, of immutable and necessary "*laws*" which rule the universe, and we cannot accept it as true simply because, in the light of our logic, it lacks all meaning....
>
> Nature will not appear ... as a monstrous and incorrigibly exact clockwork mechanism where everything that happens is what must happen because it could not but happen, and where all is foreseeable if one knows how the mechanism works.
>
> (1931a:169–70)

De Finetti returns often in his writings to this criticism of determinism and to a consideration of what should replace it. He also (De Finetti 1931a) explicitly rejects enlightenment rationalism in favour of a relativistic, and even irrational, mentality. Thus he says:

> ... the subjective theory of probability ... [is] ... an example of the application of the relativistic mentality to such an increasingly important branch of modern mathematics as the probability calculus, and as an essential part of the new vision of science which we want to give in an irrationalist, and, as we shall say, probabilist form.
>
> (1931a:172)

As we observed at the end of Chapter 2, these anti-enlightenment themes are very characteristic of the twentieth century, and perhaps especially of the 1930s when De Finetti was writing.

Although De Finetti refers to all the French authors mentioned above, his most frequent reference is to Poincaré's chapter on the calculus of probabilities in *Science and Hypothesis* (1902: Chapter XI, 183–210). Here Poincaré does indeed introduce subjective probability, which he says is the appropriate concept when a gambler is trying a single *coup* (1902:187–8). However, Poincaré goes on to argue that there is objective probability which manifests itself in a long sequence of repetitions. It looks as if De Finetti accepted Poincaré's notion of subjective probability but did not see any need for having objective probability as well. However, Poincaré has an argument for objective probability based on the insurance business. How could insurance companies make regular profits, he asks, if there was not some objective reality corresponding to their probability calculations? This argument obviously puzzled De Finetti, because he comments on it as follows:

> It seems strange that from a subjective concept there follow rules of action that fit practice. And Poincaré keeps explaining why the subjective explanation seems insufficient to him, mentioning practical applications in the field of insurance. "There are many insurance companies that apply the

> rules of the probability calculus, and they distribute to their shareholders dividends, whose objective reality is incontestable."
>
> (De Finetti 1931a:194)

Poincaré's example might be criticised in the light what happened at Lloyd's of London. This insurance company not only failed to distribute dividends, but even brought financial disaster to many of its 'names'. Is this an argument for the subjective approach to probability? Did the managers of Lloyd's formulate subjective probabilities for various events, which, although perfectly coherent, were rather unlucky? Or were they a bunch of incompetents who failed to apply the probability calculus correctly? Unfortunately, the whole matter is surrounded by great obscurity and allegations of fraud and corruption. So it is difficult to draw any definite conclusion.

We can now consider another important difference between Ramsey and De Finetti. It is to De Finetti rather than Ramsey that we should attribute the concept of exchangeability. This remark needs a little qualification since one of Ramsey's manuscript notes, published for the first time in 1991, does contain a derivation of Laplace's Rule of Succession in the special case $r = n$ using an argument quite similar to the one given above (pp. 70–3). Ramsey make the derivation under the condition: 'Suppose chance a priori of μ out of n + 1 being A is f(μ), all permutations equally probable.' (1991:278). The condition of all permutations being equally probable is equivalent in this context to De Finetti's exchangeability. Galavotti, who was the first to publish this passage, suggests that Ramsey took this condition 'from his teacher Johnson, who had introduced a 'permutation postulate' (1994:333).[10] However, we have here only a short unpublished note dealing with a very special case. This does not compare with De Finetti (1930b:121), who defined the concept explicitly,[11] and then went on to develop the mathematical theory of exchangeable random quantities in a series of important papers which culminated in his 1937. Since De Finetti wanted to eliminate objective probabilities completely in favour of subjective probabilities, he had more of a stimulus for developing the theory of exchangeability than had Ramsey, who, in his 1926 book at least, advocated, like Poincaré, a two-concept view of probability with both objective and subjective probabilities. I will return to Ramsey's two-concept view in Chapter 8, after I have given a detailed account of the two principal objective theories of probability in Chapters 5, 6 and 7.

5 The frequency theory

The frequency theory of probability was first developed in the middle of the nineteenth century by the Cambridge school of Ellis and Venn, and it can be considered as a 'British empiricist' reaction against the 'Continental rationalism' of Laplace and his followers. It became popular during another flowering of empiricism brought about by the Vienna Circle. For a while (1922–36) this twentieth-century version of empiricism had its main centre on the Continent, but, with the dispersion of the Vienna Circle, it returned to its English-speaking homelands. The frequency theory of probability was further developed at this time by two thinkers closely associated with the Vienna Circle: Hans Reichenbach and Richard Von Mises. I prefer Von Mises' version of the theory, and will expound it in what follows. Reichenbach's version is to be found in his 1949 book.

Von Mises first published account of the frequency theory is in his 1919 paper, but his most famous work on the subject is his 1928 book *Probability, Statistics and Truth.* Von Mises died in 1953. His posthumous work, *Mathematical Theory of Probability and Statistics* (1964a), assembled from his papers by his widow Hilda Geiringer, contains his final thoughts on the subject, replies to criticisms and also contains the discussion of some interesting mathematical points.

Probability theory as a science

In the logical approach probability theory is seen as a branch of logic, as an extension of deductive logic to the inductive case. In the subjective approach probability theory is seen as concerned with the degrees of belief of particular individuals. In contrast to both these views, the frequency approach sees probability theory as a mathematical science, such as mechanics, but dealing with a different range of observable phenomena. In the preface to the third German edition of his *Probability, Statistics and Truth* (1950), Von Mises characterises his theory in exactly this way. He says:

> The essentially new idea which appeared about 1919 (though it was to a certain extent anticipated by A. A. Cournot in France, John Venn in England, and Georg Helm in Germany) was to consider the theory of probability as a science of the same order as geometry or theoretical mechanics.
>
> (1950:vii)

The year 1919 was of course when Von Mises published his first paper on the frequency theory; but, even if the idea was not quite as new as he here implies, his characterisation of it seems to be quite accurate.

Concerning this alleged science of probability, we might first ask: 'what is its subject matter?' Von Mises answers as follows: '... just as the subject matter of geometry is the study of space phenomena, so probability theory deals with mass phenomena and repetitive events' (1950:vii). Von Mises' view of geometry as a science is somewhat controversial. Since, however, no one doubts that mechanics is a branch of science, it might therefore be better to state Von Mises' position as follows. Probability theory is a mathematical science like mechanics, but, instead of dealing with the motions and states of equilibrium of bodies and the forces which act on them, it treats 'problems in which either the same event repeats itself again and again, or a great number of uniform elements are involved at the same time' (Von Mises 1928:11). This emphasis on collections is in striking contrast to the subjective theory, which considers probabilities to be assigned by specific individuals to particular events. In the frequency theory, probabilities are associated with collections of events or other elements and are considered to be objective and independent of the individual who estimates them, just as the masses of bodies in mechanics are independent of the person who measures them.

Von Mises gives a number of examples of his repetitive events and mass phenomena, which can be divided into three categories. First come 'games of chance' where we deal, for example, with a long sequence of tosses of a particular coin. Second we have certain 'life' or more generally 'biological' statistics. Here we might deal with the set of German men who were 40 in 1928 or with the set of plants grown in a certain field. Lastly, we have a number of situations which occur in physics, for example the consideration of the molecules of a particular sample of gas. In all the examples cited, a particular 'attribute' occurs at each of the 'elements' which make up the set of repetitive events or mass phenomenon, but this attribute varies from one element to another. For example, on each toss of the coin 'heads' or 'tails' occurs, each of the German men either dies before reaching the age of 41 or survives into his 42nd year, the plants in the field yield a certain quantity of grain and finally each of the molecules of the gas has a certain velocity. Associated with each repetitive event or mass phenomenon, we have a set of attributes which we regard as a priori possible. These form what Von Mises calls the *attribute space.*

The attribute space, usually denoted by Ω, is one concept introduced by Von Mises which is to be found in most modern textbooks of probability theory. Unfortunately, its name has been changed from attribute space to the definitely worse *sample space.* Ω is a set of possible outcomes. Now, naturally if we take a sample, some of these outcomes will appear, but the set of possibilities has nothing essentially to do with sampling, and any given sample is unlikely to contain all the members of Ω. There is thus a case for reviving Von Mises' terminology. Strictly speaking, Ω should be said to consist of *elementary attributes,* since any subset of Ω is itself an attribute or possible outcome. Take, for example, the case of rolling a die. Here the elementary attributes are 1, 2, ..., 6, so that $\Omega = \{1, 2, ..., 6\}$.

Consider the subset of Ω, A = {2, 4, 6}. This is the (non-elementary) attribute 'even'.

Von Mises introduced the technical term *collective* to describe repetitive events or mass phenomena of the above types. More precisely he says that a collective: 'denotes a sequence of uniform events or processes which differ by certain observable attributes, say colours, numbers, or anything else' (Von Mises 1928:12). It is useful to make a distinction between *empirical collectives* and *mathematical collectives.* An empirical collective is something which actually exists in the real world and which can be observed. Examples would be a sequence of tosses of a coin carried out one Monday morning at a particular place, or the molecules of a jar of gas prepared in a particular laboratory at a particular time. It is obvious that an empirical collective has only a finite number of members. A mathematical collective, on the other hand, consists of an *infinite* sequence $\{\omega_1, \omega_2, ..., \omega_n, ...\}$ where for all n, ω_n is a member of Ω. Some problems naturally arise about the relation between the large, but finite, collections in the real world and the infinite sequences in the mathematical theory, and we will now consider these problems briefly.

The first point to note is that a mathematical collective consists of an ordered sequence, numbered 1, 2, and so on. This fits the case of coin tossing quite nicely because there is always a first toss, a second toss, etc. The other examples of collectives are not, however, naturally ordered. The plants in a field or the molecules in a gas do not occur in a particular sequence. Of course, we can number the plants in a field in some way and thus reduce them to an ordered sequence, but this can be done in a variety of different ways. In thus using an ordered sequence to represent the empirical collective of plants, we are implicitly assuming that the way in which the plants are ordered is of no importance and will not affect the results obtained. This may be true, but it is a substantial assumption. I will return to this problem in Chapter 7 in connection with the propensity theory.

Let us now consider the key question. In Von Mises' theory a finite empirical collective is represented in the mathematical theory by an infinite mathematical collective. Is this representation of the large finite by the infinite legitimate? Von Mises answers 'yes', because this is something which occurs everywhere in physics. In mechanics, for example, we have point particles to represent bodies with a size, infinitely thin lines to represent lines with a finite thickness, and so on. As a former teacher of mine was wont to say: 'in physics, "at infinity" means "on the other side of the lab".' Von Mises argues that he is trying to present probability theory as a mathematical science like mechanics, but it is unreasonable to expect him to make it more rigorous than mechanics. If the representation of the finite by the infinite is regarded as satisfactory in mechanics, it must surely be allowed in probability theory. Von Mises fully admits that infinite sequences, infinitely thin lines, etc. are mathematical abstractions from or idealisations of empirical reality, but such abstractions are necessary, he claims, in order to make the mathematical representation of reality tractable. As he says:

> Attempts have been made to construct geometries in which no 'infinitely narrow' lines exist but only those of definite width. The results were meagre because this method of treatment is much more difficult than the usual one. Moreover, a strip of definite width is only another abstraction no better than a straight line ...
>
> (Von Mises 1928:8)

These arguments of Von Mises do have a certain force, but yet they have not convinced everybody. However, it will be more convenient to return to this question of representing the large finite by the infinite, after the frequency theory has been somewhat further developed, and we shall do so in the section 'The limiting frequency definition of probability'.

The relation between empirical and mathematical collectives is part of a general view of Von Mises of how mathematical sciences relate to the empirical material with which they are concerned. This view is illustrated in Figure 5.1.

Since Von Mises was an empiricist, the starting point for him was always some observable phenomenon such as an empirical collective. To deal with such phenomena, we obtain by abstraction or idealisation some mathematical concepts, such as, in this instance, the concept of mathematical collective. We next establish on the basis of observation some empirical laws which the phenomena under study obey. Then again by abstraction or idealisation we obtain from these empirical laws the axioms of our mathematical theory. Once the mathematical theory has been set up in this way, we can deduce consequences from it by logic, and these provide predictions or explanations of further observable phenomena. In the next section we shall make a further application of this scheme to the case of probability theory by considering what empirical laws empirical collectives obey, and how these laws are established.

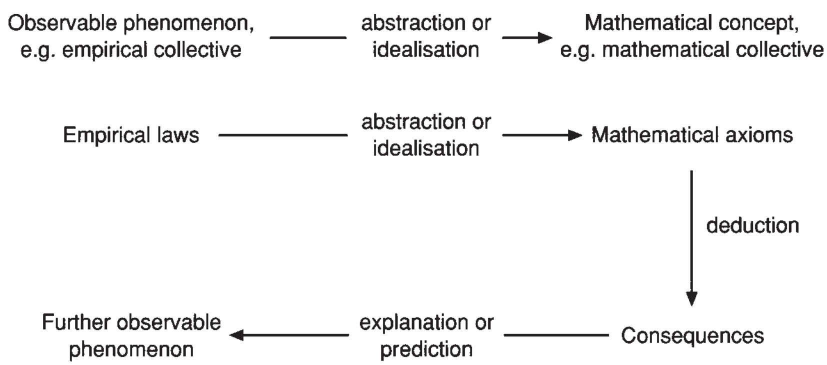


Figure 5.1 Von Mises' view of the relation between observation and theory in a mathematical science

The empirical laws of probability

There are, according to Von Mises, two empirical laws which are observed to hold for empirical collectives. He explains the first as follows:

> It is essential for the theory of probability that experience has shown that in the game of dice, as in all the other mass phenomena which we have mentioned, the relative frequencies of certain attributes become more and more stable as the number of observations is increased.
>
> (Von Mises 1928:12)

Von Mises refers to this increasing stability of statistical frequencies as: 'the "primary phenomenon" (Urphänomen) of the theory of probability.' (1928:14). I will call it the *Law of Stability of Statistical Frequencies* – a name suggested by Keynes (1921:336). Let us now attempt to state the law a little more precisely and examine some of the evidence in its favour.

Let A be an arbitrary attribute associated with a particular collective. If Ω is the attribute space of the collective, then $A \subseteq \Omega$. Suppose that in the first n members of the collective A occurs $m(A)$ times, then its relative frequency is $m(A)/n$. The Law of Stability of Statistical Frequencies is that as n increases $m(A)/n$ gets closer and closer to a fixed value. An illustration of this law is provided by Figure 5.2.

Figure 5.2 shows in graphical form the results of tossing an ordinary coin 400 times. The relative frequency (or frequency ratio) or heads is plotted against the number n of tosses. The first toss must have yielded heads, because the frequency ratio starts at 1. It then oscillates in an irregular fashion, but, after 200 or so tosses, the oscillations become less, and the frequency ratio settles down near the value of $^1/_2$.

According to Von Mises, the Law of Stability of Statistical Frequencies is confirmed by observations in all the games of chance (dice, roulette, lotteries, etc.), by insurance companies, in biological statistics, and so on. Of course, the confirming data were not in general obtained as a result of a deliberate attempt to check the law, but were collected in the course of pursuing other activities in these fields. In this connection Von Mises (1928:58–64) mentions the case of Chevalier de Méré, which we discussed earlier (pp. 3–6). In Von Mises' terminology, M. de Méré was concerned with two different collectives. The members of the first ($\mathbf{C}_1$) consisted of four throws of a single die, and the attribute (A) was getting a 6 on at least one of the throws. The members of the second collective ($\mathbf{C}_2$) consisted of twenty-four throws of two dice, and the attribute (B) was getting at least two 6s. M. de Méré discovered through empirical observation that the relative frequency of A in $\mathbf{C}_1$ tended to a value just greater that 0.5, whereas the relative frequency of B in $\mathbf{C}_2$ tended to a value just less than 0.5. He thus obtained through diligent observations two striking confirmations of the Law of Stability of Statistical Frequencies, though this was hardly his motive in making the observations. Similar considerations apply in the case of insurance companies, and so on.

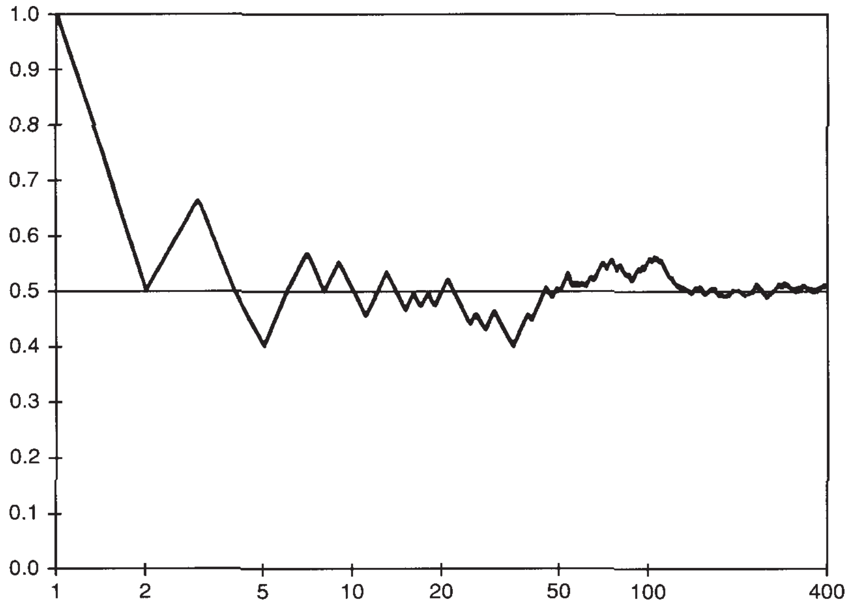


Figure 5.2 Some empirical evidence for the law of stability of statistical frequencies. The frequency ratio of heads in a sequence of tosses of a coin (logarithmic scale for the abscissa)

So far, I agree with Von Mises. There does indeed seem to be a rough empirical law of the kind he suggests, and it does appear to be confirmed by observations in a number of different areas. His next step is, however, more doubtful, since he tries to state the law in a more precise form as follows:

> If the relative frequency of heads is calculated accurately to the first decimal place, it would not be difficult to attain constancy in this first approximation. In fact, perhaps after some 500 games, this first approximation will reach the value of 0.5 and will not change afterwards. It will take us much longer to arrive at a constant value for the second approximation calculated to two decimal places.... Perhaps more than 10,000 casts will be required to show that now the second figure also ceases to change and remains equal to 0, so that the relative frequency remains constantly 0.50.
>
> (Von Mises 1928:14)

My doubt about this passage is the following. Von Mises is at this stage stating what purports to be an empirical result obtained by observation without using any theoretical or mathematical considerations. So to check the claim that 'after some 500 games, this first approximation will reach the value 0.5 and will not change afterwards', an experiment of the following kind would have to be carried out. The coin would have to be tossed say 1,000 times over and over again, and, on each repetition, it would have to be checked that from 500 onwards the relative frequency of heads was 0.5 (to one decimal place). But what then about the claim that 'perhaps more than 10,000 casts will be required to show that now the second figure also ceases to change and remains equal to 0'? Here we would have to toss the coin say 11,000 times over and over again, and, on each repetition, check that, after 10,000 tosses, the relative frequency of heads was 0.50 (to two decimal places). I have carried out some coin tossing, and will give the results I obtained in Chapter 7. However, I discovered that to toss a coin even 2,000 times is a long and tedious business. Thus, to toss a coin 11,000 times and then repeat the experiment over and over again would be formidably dull and time-consuming. Moreover, a figure of the order of 10,000 can be obtained in a few lines by a very simple mathematical calculation, as I will now show.

Suppose a coin for which prob(heads) = $^1/_2$ is tossed n times, and heads is obtained m times. Suppose further that the tosses are independent. Then by a classic result of De Moivre (see note 4 to Chapter 1, pp. 206–7), m/n is approximately normally distributed. More precisely,

$$\frac{m/n-0.5}{0.5/\sqrt{n}}$$

is for large n approximately normally distributed with zero mean and unit standard deviation. Thus, from tables of the normal distribution it follows that with 95 per cent probability, we have

$$\left|\frac{m}{n}-0.5\right| \leq \frac{0.98}{\sqrt{n}} \tag{5.1}$$

So we get that the following results hold with 95 per cent probability. If $n = 500$, m/n lies in the interval [0.456, 0.544]. This should indeed give constancy to one decimal place as Von Mises claims. If $n = 10{,}000$, m/n lies in the interval [0.4902, 0.5098], which does not quite give constancy to two decimal places. However for $n = 50{,}000$, m/n lies in the interval [0.4956, 0.5044], which should give constancy to two decimal places. Moreover Equation 5.1 above tells us that, in general, m/n will approach its limiting value of 0.5 at the rate of $1/\sqrt{n}$.

These calculations suggest a rather different relationship between theory and observation from that claimed by Von Mises (see Figure 5.1 above). According to Von Mises, an empirical law is obtained by observation, and a mathematical axiom of the theory is abstracted from it. Now a rough empirical law might indeed be obtained directly from observation, but, to make it more precise, it looks as if we should temporarily abandon observation in favour of mathematics. Mathematical calculations suggest more precise versions of the empirical law, e.g. that the frequency is likely to remain constant to one decimal place after 500 goes, and that, in general, the frequency is likely to converge to its limit at the rate of $1/\sqrt{n}$. These results could then be checked by further observations. In short, there seems to be more of a two-way interaction between observations and theory than Von Mises suggests. We will see in Chapter 7 that this interactive process is captured better in the propensity than the frequency theory, but now let us return to the development of Von Mises' theory.

The first law of empirical collectives was fairly well known before Von Mises. The second law is, however, original to him. Indeed, he considers its formulation to be one of his major advances. Speaking of the efforts of his predecessors in the scientific tradition (Venn and others), he says: 'These attempts, ..., did not lead, and could not lead, to a complete theory of probability, because they failed to realise one decisive feature of a collective ...' (Von Mises 1928:22). This feature of the empirical collective is its lack of order, that is its *randomness.*

Von Mises' treatment of randomness is indeed one of the most interesting and original parts of his theory. Von Mises (1928:23) begins by considering the following simple example. Suppose we are walking down a road at the side of which there are large stones at intervals of a mile, and small stones between them at intervals of $^1/_{10}$ of a mile. The first empirical law is certainly satisfied, because the attribute large stone has a limiting frequency of $^1/_{10}$, and the attribute small stone a limiting frequency of $^9/_{10}$. Yet Von Mises does not think that this is a genuine collective, because the sequence of results is perfectly determined. After a large stone, we know that the next stone will be small, and so on. This is in complete contrast with the examples of empirical collectives so far given. For example, in coin tossing whatever sequence of heads and tails has been observed so far, we have no idea what the result of the next toss will be, and similarly in the other examples. Thus genuine empirical collectives are disordered, i.e. satisfy some law of randomness. But how can we formulate this law?

Von Mises' ingenious idea is that we should relate randomness to the failure of gambling systems. A gambling system in, for example, roulette is something of the following kind: 'Bet on red after a run of three blacks', or 'Bet on every seventh go', etc. Undoubtedly, over a long period of time, many different such gambling systems have been tried out. However, as Von Mises says:

> The authors of such systems have all, sooner or later, had the sad experience of finding out that no system is able to improve their chances of winning in the long run, i.e., to affect the relative frequencies with which different colours or numbers appear in a sequence selected from the total sequence of the game.
>
> (1928:25)

In other words, not only do the relative frequencies stabilise around particular values, but these values remain the same if we choose, according to some rule, a subsequence of our original (finite) sequence. Let us call this second empirical law the *Law of Excluded Gambling Systems.* Von Mises now makes a most suggestive comparison:

> An analogy presents itself at this point which I shall briefly discuss. The system fanatics of Monte Carlo show an obvious likeness to another class of

> 'inventors' whose useless labour we have been accustomed to consider with a certain compassion, namely, the ancient and undying family of constructors of 'perpetual-motion' machines.
>
> (1928:25–6)

The failure of all attempts to construct a perpetual-motion machine provided excellent evidence for the Law of Conservation of Energy. Indeed we could describe that law as the Law of Excluded Perpetual-motion Machines. In just the same way the failure of gambling systems provides excellent evidence for our empirical law of randomness.

Despite the very strong empirical evidence to which Von Mises alludes, the idea of a successful gambling system continues to haunt the minds of compulsive gamblers. Dostoyevsky, himself a compulsive gambler, has portrayed their psychology with great brilliance in his novel *The Gambler.* Here is a passage in which the hero of the novel describes some of his thoughts at the roulette table:

> But on the other hand I drew one conclusion, which I think is correct: in a series of pure chances there really does exist, if not a system, at any rate a sort of sequence – which is, of course, very odd. For example, it may happen that after the twelve middle numbers, the last twelve turn up; the ball lodges in the last twelve numbers twice, say, and then passes to the first twelve. Having fallen into the first twelve it passes again to the middle twelve, falls there three or four times running, and again passes to the last twelve, and from there, again after two coups, falls once more into the first twelve, lodges there once and then again falls three times on the middle numbers, and this goes on for an hour and a half or two hours: one, three, two; one, three, two. This is very entertaining. One day, or one morning, it will happen, for example, that red and black alternate, changing every minute almost without any order, so that neither red nor black ever turns up more than two or three times in succession. The next day, or the next evening, red only will come up many times running, twenty or more, for example, and go on doing so unfailingly for a certain time, perhaps during a whole day.
>
> (1866:38–9)

We see that Dostoyevsky's hero, and presumably Dostoyevsky himself, believed that 'there really does exist, if not a system, at any rate a sort of sequence'. The illusory nature of this belief is well demonstrated by the fact that Dostoyevsky continually lost money at the roulette table – much to his wife's sorrow. A much more rational attitude was displayed by the multi-millionaire John Paul Getty, who when asked on a television interview whether he ever gambled replied: 'If I wanted to gamble, I would buy a casino.'

The limiting frequency definition of probability

We have now introduced the two empirical laws of probability and argued that they are well supported by the observations of le Chevalier de Méré and

Dostoyevsky, as well as by those of the more disciplined employees of insurance companies and of meticulous research scientists conducting agricultural trials. The next step in Von Mises' programme is to obtain the axioms of the mathematical theory by abstraction (or idealisation) from these empirical laws. These axioms apply of course to mathematical collectives of the form $\mathbf{C} = \{\omega_1, \omega_2, ..., \omega_n, ...\}$, where for all n ω_n is a member of Ω, the attribute space. It is easy to obtain the first axiom from the Law of Stability of Statistical Frequencies. It may be stated as follows:

Axiom of convergence

Let A be an arbitrary attribute of a collective $\mathbf{C}$, then $\lim_{n \to \infty} m(\mathrm{A})/n$ exists

We now *define* the probability of A in $\mathbf{C}$ [$\mathrm{P}(\mathrm{A} \mid \mathbf{C})$] as $\lim_{n \to \infty} m(\mathrm{A})/n$. This is the famous limiting frequency definition of probability. It is worth noting that this definition makes all probabilities conditional, and that this is one of the few points in common between the frequency and the logical theories. Yet even here there is a difference. In the logical theory, the probability of a hypothesis is always conditional on some body of evidence. Similarly, in the subjective theory, the probability of an event is always conditional on the individual who is assigning a betting quotient, and hence indirectly on the set of beliefs of that individual. In the frequency theory all probabilities are conditional, but they are conditional not on evidence or a set of beliefs, but on a particular collective of which the particular attribute in question is taken as one of the outcomes. This important difference between a body of evidence or belief on the one hand and a collective on the other actually constitutes a way of characterising the difference between epistemological and objective interpretations as I will try to show in Chapter 7.

Having introduced the limiting frequency definition of probability, let us now examine some criticisms of it. One of the main objections to the theory is that it is too narrow, for there are many important situations where we use probability but in which nothing like an empirical collective can be defined. As Keynes puts it, speaking of an earlier version of the frequency theory: 'Part of the plausibility of Venn's theory is derived, I think, from a failure to recognise the narrow limits of its applicability' (1921:96). This alleged disadvantage is, however, considered by Von Mises to be a strong point in favour of his theory. He states clearly that 'Our probability theory has nothing to do with questions such as: "Is there a probability of Germany being at some time in the future involved in a war with Liberia?"' (Von Mises 1928:9). We can only, he claims, introduce probabilities in a mathematical or quantitative sense where there is a large set of uniform events, and he urges us to observe his maxim: 'FIRST THE COLLECTIVE – THEN THE PROBABILITY' (Von Mises 1928:18).

There is indeed something to be said for Von Mises' desire to limit the scope of the mathematical theory. The history of probability affords some curious examples of 'numerical' probabilities. Todhunter (1865: 408–9), for example, records the following evaluations carried out by the eighteenth-century probabilist Condorcet. The probability that the whole duration of the reigns of the seven kings of Rome was 257 years was reckoned by him to be 0.000792, whereas the probability that it was 140 years came to 0.008887. He also calculated the probability that the augur Accius Naevius cut a stone with a razor. This came to the more rounded figure of 10^{-6}.

Von Mises' view on this point is connected with some general theories of his about the evolution of science. Von Mises (1928:1) quotes with approval Lichtenberg's maxim that 'All our philosophy is a correction of the common usage of words.' We can, according to Von Mises, *start* with the imprecise concepts of ordinary language but when we are constructing a scientific theory we must replace these by more precise concepts. Further, he thinks that these precise concepts should be introduced *by means of explicit definitions.* I will call this *Von Mises' definitional thesis.* The example he cites in this context is the mechanical concept of work. Of course we use the word 'work' in a variety of ways in ordinary language, but in mechanics we define work as force times distance or more precisely we set

$$W_{\mathrm{a}}^{\mathrm{b}} = \int_{\mathrm{a}}^{\mathrm{b}} \mathbf{F}.\mathrm{d}\mathbf{s}$$

where $W_{\mathrm{a}}^{\mathrm{b}}$ is the work done in moving from a to b in a conservative force field **F(x).** Many things ordinarily counted as work are excluded by this definition, e.g. the work involved in writing a book, the work of holding steady a heavy tray of sandwiches so that guests can help themselves, etc. The vague concept of ordinary language has been delimited and made precise by a definition.

Exactly the same applies, according to Von Mises, in the case of probability. We can of course start with the vague ordinary language concept of probability, but for scientific purposes it must be made precise by a definition. This is done by the limiting frequency definition of probability. This definition excludes some ordinary language uses of probability for which a collective cannot be defined, but this is no bad thing. On the contrary, it is positively beneficial to exclude some vague uses of probability which are unsuitable for mathematical treatment. Von Mises sums up this line of argument as follows:

> 'The probability of winning a battle', for instance, has no place in our theory of probability, because we cannot think of a collective to which it belongs. The theory of probability cannot be applied to this problem any more than the physical concept of work can be applied to the calculation of the 'work' done by an actor in reciting his part in a play.
>
> (Von Mises 1928:15)

There was much to be said for this view of Von Mises when he formulated it in 1928. At that time the only method of evaluating probabilities apart form using observed frequencies involved the Principle of Indifference, and that principle was known to lead to insoluble paradoxes. In the years 1930–1, however, the first papers of the new subjective approach were published, and these gave a method for measuring probabilities *qua* degrees of belief in such a way that the axioms of probability could be derived from the plausible condition of coherence. These new results showed that it was possible to extend quantitative probabilities and the mathematical calculus to many cases where no collective was involved. Let us consider the two cases which Von Mises himself cites. It is not possible to bet on whether Germany will be at some time in the future involved in a war with Liberia, for such a bet might never be settled. However, we have only to change the example to that of the probability that Germany will in the next fifty years be involved in a war with Liberia, and a subjective probability can be introduced in the standard fashion. If a battle is due to begin tomorrow, we can certainly bet on its outcome, and so Von Mises' second example falls immediately within the domain of the subjective theory. It is not that Von Mises' frequency theory has been shown to be wrong, but only that it is possible, using the methods of the subjectivists, to extend the mathematical calculus to examples which lie outside the scope of the frequency theory.

Another related criticism of the frequency theory is that it does not deal with the rôle of probability in induction and confirmation. This objection is made by De Finetti in his article on Von Mises: 'If an essential philosophical value is attributed to probability theory, it can only be by assigning to it the task of deepening, explaining or justifying the reasoning by induction. This is not done by Von Mises, ...' (De Finetti 1936:361). Once again Von Mises replies by agreeing that such is indeed a consequence of his theory. As he says in his 1950 preface to the third German edition of his 1928 book:

> According to the basic viewpoint of this book, the theory of probability in its application to reality is itself an inductive science; its results and formulas cannot serve to found the inductive process as such, much less to provide numerical values for the plausibility of any other branch of inductive science, say the general theory of relativity.
>
> (1950:ix)

Naturally, the important questions of induction and confirmation need to be discussed, but it does not necessarily follow that the mathematical calculus of probability is the correct tool for dealing with these problems. It could be, for example, that judgements about the confirmation of a hypothesis by evidence are inherently qualitative rather than quantitative in nature.

Let us now examine Von Mises' definitional thesis that all the concepts of a precise mathematical science should be introduced by explicit definitions. His example of the concept of work in mechanics does indeed show that some concepts are introduced in this way, but does this apply to all the concepts of a mathematical theory? After all, if we define one concept it must be in terms of others. Thus, the concept of work is defined in terms of those of force and distance. If therefore we demand that all concepts be defined, will we not be led either to an infinite regress or a vicious circle? Moreover, there seems to be another way of introducing concepts apart from that of explicit definitions. For example, we could develop Newtonian mechanics by taking the concepts of *force* and *mass* as primitive and characterising them by a set of axioms. Similarly, we might take *probability* as a primitive notion, and characterise it by the axioms of the theory. Cramér (1946) adopts an approach of this sort and criticises Von Mises' explicit definition of probability as follows:

> ... some authors try to introduce a system of axioms directly based on the properties of frequency ratios. The chief exponent of this school is Von Mises ..., who defines the probability of an event as the *limit of the frequency* v/n of that event, as n tends to infinity. The existence of this limit, in a strictly mathematical sense, is postulated as the first axiom of the theory. Though undoubtedly a definition of this type seems at first sight very attractive, it involves certain mathematical difficulties which deprive it of a good deal of its apparent simplicity. Besides, the probability definition thus proposed would involve a mixture of empirical and theoretical elements, which is usually avoided in modern axiomatic theories. It would, e.g., be comparable to defining a geometrical point as the limit of a chalk spot of infinitely decreasing dimensions, which is usually not done in modern axiomatic geometry.
>
> (1946:150)

It is interesting to note that Russell (1914: 119–20) in his book *Our Knowledge of the External World* did propose to define points in a way not very different from the one Cramér describes here. On the other hand, Cramér is correct that most modern treatments of geometry following Hilbert's 1899 *Foundations of Geometry* do introduce point as a primitive undefined notion which is characterised axiomatically.

Von Mises might not object to the idea that in a mathematical science there are some basic notions, such as force and mass in the case of mechanics or point and line in the case of geometry, and that the other notions of the science, such as work in the case of mechanics or quadrilateral in the case of geometry, are defined in terms of these basic notions. However, I am sure he would add that, if the theory is to be a branch of empirical science and not just pure mathematics, these basic notions need to be given operational definitions in terms of observables. Von Mises gives a clear statement of his operationalism in the 1950 preface to the third German edition of 1928, where he writes: 'The relative frequency of the repetition is the 'measure' of probability, just as the length of a column of mercury is the 'measure' of temperature.' (1950:viii).

Von Mises derived his operationalist/positivist ideas from Mach, whom he greatly admired. After giving his views on the need for defining probability, Von Mises adds: 'The best information concerning ... the general problem of the formation of concepts in exact science can be found in E. MACH .... The point of view represented in this book corresponds essentially to MACH's ideas.' (1928:Footnote 7, 225). Von Mises (1938) later gave a glowing account of Mach's philosophy in his article 'Ernst Mach und die empiristische Wissenschaftsauffassung', and wrote in a summary of his book on positivism: 'The author is a devoted disciple of *Mach*' (1940:524). These tributes are indeed appropriate, for Von Mises in his development of probability theory follows exactly the pattern of Mach's development of mechanics.

In his *Science of Mechanics,* Mach criticises earlier treatments of Newtonian mechanics for failing to give an adequate account of the concept of mass, and he attempts to remedy this defect by proposing an operational definition of mass in terms of observables (Mach 1883:264–71, 298–305). Mach gives three experimental propositions which are supposed to be established by observations and then bases his definitions of mass and force on these. It seems to me evident that Von Mises modelled his account of probability on this account of mechanics, since he first introduces the Law of Stability of Statistical Frequencies, which is supposed to be established by observations, and then bases his definition of probability on this law.[1]

Mach's positivism and operationalism have been much criticised, and nowadays most philosophers of science prefer a rather different account of the relationship between observation and the theoretical concepts of natural science. It is no longer widely believed that theoretical concepts should be directly defined in terms of observables, and it is more generally held instead that such concepts should be initially be undefined and then connected to experience in a more indirect fashion. In Chapter 7, I will present a non-operationalist account of how the theoretical concepts of natural science can be linked to observation and experiment, using as illustration Mach's example of Newtonian mass. I will then show how this non-operationalist approach leads to a different account of probability from Von Mises'. This new account is one version of the propensity theory.

The limiting frequency definition of probability is supposed to be an operational definition of a theoretical concept (probability) in terms of an observable concept (frequency). However, it could be claimed that it fails to provide a connection between observation and theory because of the use of limits in an infinite sequence. It is well known that two sequences can agree at the first n places for any finite n however large and yet converge to quite different limits. Suppose I toss a coin 1,000 times and the observed frequency of heads is approximately 1/2. This is quite compatible with the limit being quite different from 1/2. Therefore, it is argued, Von Mises' definition fails to link theory and observation.

We have met a very similar objection earlier when we considered the question of whether finite empirical collectives could be represented by the infinite sequences of mathematical collectives. Von Mises' answer to the difficulty was to say that such representations of the finite by the infinite occur everywhere in mathematical physics, and his aim was only to present probability theory in a fashion which was as rigorous as the rest of mathematical physics. He surely could not hope to make it more rigorous. This point of view can be illustrated by comparing Von Mises' limiting frequency definition of probability with a typical use of limits in mathematical physics. For this purpose let us consider how the density at a point in a fluid is defined. This is an example highly appropriate for Von Mises, who, in addition to his work on probability, made important contributions to fluid mechanics, and indeed he mentions a related example (Von Mises 1928:84–5).

The definition of the density ρ at a point P in a fluid is illustrated in Figure 5.3. As shown in Figure 5.3(a) we take a small volume δV around P, and suppose it contains a mass δM. We then define the density ρ at P as the limit as $\delta V \to 0$ of $\delta M/\delta V$. This seems exactly parallel to Von Mises' limiting frequency definition of probability except that here we have a quantity growing smaller and smaller, instead of one growing larger and larger. However, it might be argued that the situation in fluid mechanics is in some respects worse, because we know that fluids are not continuous but are actually composed of molecules. Consequently when δV is sufficiently small to be comparable with the mean free path of a molecule, the values of δM will fluctuate violently with the random fluctuations of the molecules. Thus, if we really could take a series of readings of δM for successively smaller values of dV, the result would appear somewhat as shown in Figure 5.3(b). In the section BC, $\delta M/\delta V$ would indeed appear to be converging to a definite value, but, as δV got smaller still and entered the region AB where it was comparable with the mean free path of a molecule, $\delta M/\delta V$ would start to oscillate in an irregular fashion and all tendency towards a definite limit would be lost. This is not a purely academic point because continuous fluid mechanics does indeed break down in circumstances in which the molecular structure of matter is significant. This is the case for gases at very low pressures, for example high in the atmosphere. More exact calculations reveal that continuous fluid mechanics cannot be applied more than 200 km above the Earth's surface. Rather different considerations (involving molecular

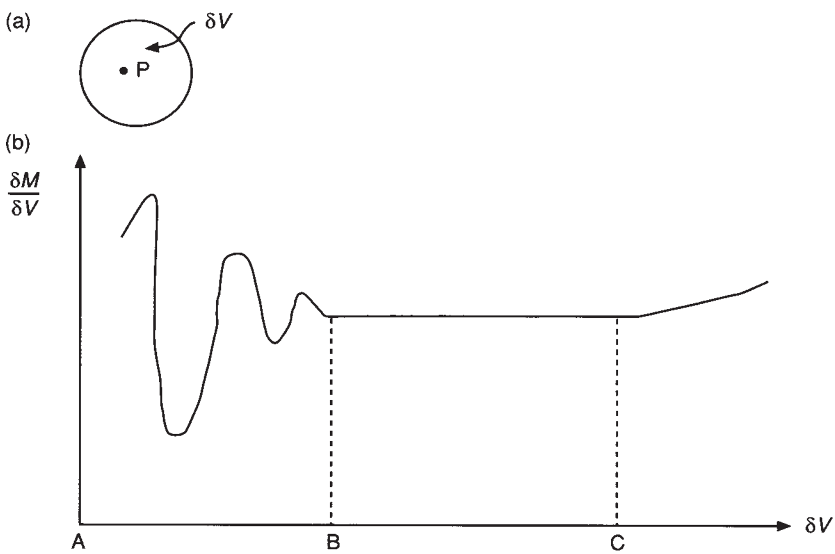


Figure 5.3 Definition of the density ρ at a point P in a fluid

interactions) show that the continuity assumptions will break down in the case of shock waves.

The situation as regards limits in fluid mechanics thus appears to be *worse* than in the frequency theory of probability. We cannot toss a coin infinitely often, but, so far as we know, m(heads)/n will continue to converge for as long as we do toss it, whereas there is a lower bound to the size of the volumes δV which we can use in taking the limit of $\delta M/\delta V$. Admittedly, there may be a corresponding difficulty in the case of the frequency theory of probability. The coin will no doubt gradually wear away as we continue to toss it, and this could alter the value of prob(heads). It remains true, however, that limits are used and regarded as quite unproblematic in fluid mechanics, and the situation as regards the limits in the frequency theory of probability does appear to be no worse, and perhaps even better. So Von Mises concludes:

> ... the results of a theory based on the notion of the infinite collective can be applied to finite sequences of observations in a way which is not logically definable, but is nevertheless sufficiently exact in practice. The relation of theory to observation is in this case essentially the same as in all other physical sciences.
>
> (1928:85)

This argument of Von Mises is a strong one, but De Finetti maintains nonetheless that there is a difference between probability theory and other physical sciences in this respect. He writes:

> It is often thought that these objections may be escaped by observing that the impossibility of making the relations between probabilities and frequencies precise is analogous to the practical impossibility that is encountered in all the experimental sciences of relating exactly the abstract notions of the theory and the empirical realities. The analogy is, in my view, illusory: in the other sciences one has a theory which asserts and predicts with certainty and exactitude what would happen if the theory were completely exact; in the calculus of probability it is the theory itself which obliges us to admit the possibility of all frequencies. In the other sciences the uncertainty flows indeed from the imperfect connection between the theory and the facts; in our case, on the contrary, it does not have its origin in this link, but in the body of the theory itself ...
>
> (De Finetti 1937:117)

What De Finetti means here can be explained by considering again the example of continuous fluid mechanics. Suppose we construct a model using that body of theory of how water behaves in a particular situation. Suppose we can fix the values of the parameters in this model empirically and can solve the equations.

Then our model will tell us exactly how the water should behave. Of course, the model will, for a variety of reasons, only be an approximation to what happens, but what the model tells us is nonetheless precise. The model is, as De Finetti says, 'a theory which asserts and predicts with certainty and exactitude what would happen if the theory were completely exact.'

Now let us contrast this with probability theory. Suppose we model the simplest case of tossing a symmetrical coin, by assuming that the tosses are independent with prob(heads) = $^1/_2$. We can then deduce (Equation 5.2) that, if there are *m* heads in *n* tosses,

$$\left|\frac{m}{n} - 0.5\right| \leq \frac{0.98}{\sqrt{n}}$$

holds *with 95 per cent probability.* The key point is that we cannot conclude from the model that the relationship holds with certainty, but only that it holds with 95 per cent probability. Indeed, all values of in the interval [0, 1] are possible, even though the probability of a value which diverges considerably from 0.5 will be very low. As De Finetti says: 'in the calculus of probability it is the theory itself which obliges us to admit the possibility of all frequencies.'

In my view, De Finetti does succeed here in showing that there is a disanalogy between probability theory and other branches of physics as regards the relationship between 'the abstract notions of the theory and the empirical realities.' In the case of probability, some extra assumptions are needed to link theory with reality. In Chapter 7 I will suggest that what is required here is a *falsifying rule for probability statements.*

In this section, it was shown how the mathematical axiom of convergence could be obtained from the empirical Law of the Stability of Statistical Frequencies, and how the axiom of convergence led to the limiting frequency definition of probability. This definition gives rise to a whole series of philosophical problems, and we have spent most of this section discussing these. The derivation of the mathematical axiom at the beginning posed no serious problem. To complete Von Mises' programme, however, we must now examine how the second mathematical axiom (the axiom of randomness) can be obtained from the empirical Law of Excluded Gambling Systems. It turns out that the formulation of the axiom of randomness does involve very considerable mathematical difficulties. These difficulties were overcome, but only by some quite subtle mathematical developments. I will deal with these matters in the next section, 'The problem of randomness*'. The final section, 'The relation between Von Mises' axioms and the Kolmogorov axioms*', will raise the same question for the frequency theory which we raised for the subjective theory, namely how the axioms relate to the standard Kolmogorov axioms. Once again we shall find some points of difference. These last two sections of this chapter have accordingly a primarily mathematical interest. We have so far dealt with most of the philosophical problems of the frequency theory which constitute the background to the propensity theory, and so the reader uninterested in the further mathematical questions can proceed directly to Chapter 6.

The problem of randomness*

The empirical Law of Excluded Gambling Systems states roughly that it is impossible to improve one's chances of winning by using a gambling system. Our problem now is to formulate a version of this for mathematical collectives which will constitute the second axiom of the mathematical theory – the axiom of randomness. To see the difficulties involved in this task, let us first formulate a 'naive' version of the axiom, which does not in fact work. Let us take **C** to be a mathematical collective with attribute space Ω, and let us suppose that **C** satisfies the first axiom (the axiom of convergence). We then have for any attribute A, where $A \subseteq \Omega$, the probability of A in **C** [Prob(A | **C**)] = $\lim_{n \to \infty} m(A)/n$. Let us further define a *place selection* or *gambling system* as a rule for selecting a subsequence **C′** of **C.** The gambling system can be said to be successful if the limiting frequency of $m(A)/n$ in **C′** (p' say) differs from its value in **C**, i.e. Prob(A | **C**). It does not matter whether p' is greater or less than Prob(A | **C**), provided there is a difference; for if p' > Prob(A | **C**), we bet on A occurring on members of the subsequence **C′**, whereas if p' < Prob(A | **C**), we bet against A occurring on members of the subsequence. In the light of these definitions, we can formulate our 'naive' (and erroneous) version of the axiom of randomness as follows. In any subsequence **C′** obtained from the original collective **C** by means of a place selection, $m(A)/n$ must continue to converge to its original value in **C**, i.e. Prob(A | **C**).

The trouble with this 'naive' axiom is that it renders the class of collectives empty except in the trivial case when the probability of each attribute is either 0 or 1. For suppose that attribute A has a probability greater than zero and less than one. By the first condition, A must appear an infinite number of times. Thus, we can choose a subsequence consisting just of attributes A. For this subsequence we have $\lim_{n \to \infty} m(A)/n = 1 \neq P(A \mid \mathbf{C})$, and so our naive axiom of randomness does not hold. Clearly, we have to restrict the class of allowable place selections or gambling systems in some way in order to avoid this unpleasant consequence. The problem is how to do this.

Von Mises himself suggested that we make the following stipulation: 'the question whether or not a certain member of the original sequence belongs to the selected partial sequence should be settled *independently of the result* of the corresponding observation, i.e., before anything is known about this result.' (Von Mises 1928:25). Undoubtedly, this is very reasonable as far as the actual betting situation is concerned. Casinos do not allow one to decide whether to bet on a particular turn of the wheel after the result is known. However, we are concerned now not with practical procedures relating to empirical collectives but with mathematical definitions relating to mathematical collectives. In formulating a mathematical definition, we have to use mathematical concepts and cannot bring in considerations about whether someone does or does not know the values of the first n members of a particular collective. Indeed, as we have seen, Von Mises himself always stressed the need to separate the mathematical from the empirical. Moreover, there was in the present instance another factor which made a precise mathematical formulation of the axiom of randomness desirable.

An objection was made to Von Mises' theory that any adequate formulation of the axiom of randomness would contradict the axiom of convergence and thus render the theory inconsistent. To show that this objection was invalid, it was highly desirable to prove that the two axioms were consistent, and, to supply such a consistency proof, a precise mathematical formulation of the axiom of randomness, using only strictly mathematical concepts, was needed.

The objection just mentioned was put forward by Fry (1928:88–91) and Cantelli (1935:§§7, 10, 12). Von Mises shows from his axiom of randomness that in any collective **C** for which the axiom holds, the binomial formula holds. So if A is any attribute for which $P(A \mid \mathbf{C}) = p$, then the probability of getting A m times on any n members of C is given by ${}^{n}C_{m}\, p^{m}(1 - p)^{n - m}$. In other words, randomness implies independence. But, according to Fry and Cantelli, this independence contradicts the axiom of convergence. Their argument is this. Using the axiom of convergence, we have $P(A \mid \mathbf{C}) = p = \lim_{n \to \infty} m(A)/n$, where as usual we suppose that A occurs $m(A)$ times in the first n members of **C**. So, given e > 0, there must be an N such that the difference between p and $m(A)/n$ is less than e for all $n > N$. But let us now consider any finite segment of the sequence immediately following the first N elements (say the elements $N + 1, N + 2, ..., N + r$). According to the binomial formula, there is a finite probability of getting A at each of these elements, namely p^r. If we get a run of such successes long enough, $m(A)/n$ will diverge from p by more than e. There is thus for any N a finite probability of such a divergence, contrary to the requirements of the limit definition.

To resolve this difficulty we need only consider the meaning of the assertion that there is a probability p^r of getting A at the $N + 1, ..., N + r$ places of **C**. According to Von Mises, the meaning is this. If we produce an infinite sequence of collectives $\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, ..., C^{(i)}, ...$ in the same way as **C**, the limiting frequency of those which have A at the $N + 1, ..., N + r$ places will be p^r. This is not at all incompatible with the collective **C** quite definitely *not* having A at all these places. We would only get a contradiction if we postulated not only that the relative frequency of A converged to p for each $\mathbf{C}^{(i)}$, but also that the convergence was uniform over the $\mathbf{C}^{(i)}$.

This I think satisfactorily solves the difficulty, but some doubts may still remain whether it is really possible to formulate mathematically an axiom of randomness which is consistent with the axiom of convergence. In the period 1919–40 these problems connected with randomness attracted a great deal of attention. Von Mises' frequency theory of probability was very popular with the Vienna Circle, and members of the circle as well as thinkers having links to the circle devoted attention to the question. A complete list of those who made contributions would include the names of Church, Copeland, Dörge, Feller, Kamke, Von Mises, Popper, Reichenbach, Tornier, Waismann and Wald. I will not, however, attempt a detailed history, but rather concentrate on the work of Wald and Church, whose combined efforts produced, in my opinion, a complete and satisfactory solution to the original problem.

In expounding the results of Wald and Church, I will, for simplicity, confine myself to mathematical collectives with attribute space {0, 1}, i.e. infinite sequences of 0s and 1s. The results can of course be generalised to collectives with other kinds of attribute space. Wald's results are contained in his 1937 paper *Die Widerspruchsfreiheit des Kollektivsbegriffes* (*The Consistency of the Concept of Collective*). The same results (but without proofs) were given in a shorter paper with the same title in 1938. This paper is reprinted in the original German in Wald's selected papers, and so is more accessible. There is also a summary in English of his work on this problem by Von Mises (1964a:39–43).

Wald's approach is not to try to define a specific allowable class of place selections or gambling systems, but rather to examine the effect of choosing this class is different ways. His main theorem is the following. If we confine ourselves to a denumerable class of place selections or gambling systems, then there exist a continuum infinity of collectives or random sequences having any assigned probability distribution. So, far from random sequences being rare or even non-existent, they are much more numerous than sequences which exhibit regularity. This result still leaves a certain arbitrariness in the choice of the class of place selections or gambling systems, but Wald tries to mitigate this by two considerations. First of all in any particular problem we certainly will not want to consider more than a denumerable set of gambling systems. Second, let us suppose that we are formulating our theory within some logical system, e.g. (to quote his example) Russell and Whitehead's *Principia Mathematica.* Within such a system we only have a denumerable set of formulas, and so can only define a denumerable set of mathematical rules.

This last remark of Wald's may have suggested to Church a way in which the class of allowable gambling systems could be specified more precisely. Church had at his disposal a mathematical theory which had recently been developed by a number of people including himself quite independently of any questions in probability theory. This was the theory of *recursive functions.* Let us define a *computable function* as a function from natural numbers to natural numbers whose value for any particular input can be computed in a finite time using some purely mechanical method which is laid down in advance. Of course, this is only an informal explanation of the concept, and the question arises whether it can be characterised more precisely. The class of recursive functions had been defined in a mathematically exact fashion, and Church (1936) suggested that we could identify computable functions with recursive functions. This is the famous *Church's thesis.* Evidence for its truth soon mounted since all other ways of explicating computable functions based on many different approaches such as λ-definability, Turing machines, Post processes, Markov algorithms, and so on, turned out to be provably equivalent to recursive functions. In his short paper 'On the Concept of a Random Sequence' (1940), Church applied these new developments to the problem which had arisen from Von Mises' frequency theory. After stating the familiar objection to the existence of collectives, he goes on to say: 'Thus a *Spielsystem* [i.e. gambling system] should be represented mathematically, not as a function, or even as a definition of a function, but as an effective algorithm for the calculation of the values of a function.' (1940:133).

This point once made must surely be recognised as correct. A gambling system after all is nothing but a rule telling us at each go whether to bet or not. Such a rule must deliver its instructions in a finite time; in other words, it must be an effective procedure for determining whether we are to bet or not. Indeed, we can think of any gambling system as a kind of miniature computer which the gambler carries with him. He feeds into the computer the number n of the go, and the results of the previous n - 1 goes. The computer then outputs an instruction whether he should bet on a particular attribute at that go. This corresponds very closely to our intuitive idea of a gambling system, and shows how this idea can be explicated in terms of computable functions. So, if we accept Church's thesis, we can define gambling systems in terms of recursive functions. Church does so as follows.

Let our original collective be $\{a_1, a_2, ..., a_n, ...\}$, where we assume a_n is 0 or 1 for all n. We can also represent a gambling system by an infinite sequence of 0s and 1s $\{c_1, c_2, ..., c_n, ...\}$ say, where $c_n = 1$ means select a_n, and $c_n = 0$ means reject a_n. We shall say that $\{c_1, c_2, ..., c_n, ...\}$ is a *recursive gambling system* if $c_n = f(b_n)$ where

1 $b_1 = 1, b_{n+1} = 2b_n + a_n$,
2 ϕ is a recursive function of positive integers;

and if the integers n are such that $c_n = 1$ are infinite in number.

(The introduction of the b_n in 1 is merely a device to ensure that our decision whether or not to choose a_n say can depend on the preceding members of the a-sequence as well as on n.)

We can now formulate the axiom of randomness as follows:

> Axiom of randomness: Let **C** be a collective to which the axiom of convergence applies. Let A be an arbitrary attribute of **C** for which $P(A \mid \mathbf{C}) = \lim_{n \to \infty} m(A)/n = p$. Let $\mathbf{C}'$ be a subsequence of **C** chosen by a recursive gambling system. Then in $\mathbf{C}'$ $\lim_{n \to \infty} m(A)/n$ exists and equals p.

Since there are only a denumerable number of recursive gambling systems, it follows from Wald's theorem that there exists a continuum infinity of collectives with any assigned probability distribution satisfying the axioms of convergence and randomness as defined above.

The work of Wald and Church thus gives Von Mises' theory a rigorous mathematical foundation. It provides a very plausible explication of Von Mises' intuitive notion of a gambling system, and, with this explication, the axioms can be formulated and proved to be consistent. Despite this success a curious problem remains which is mentioned by Church.

The above proofs of the existence of random sequences are completely valid within ordinary classical mathematics, including the basic ideas of Cantorian set theory. But classical mathematics has been criticised by the constructivists, and the above proofs do not hold in at least some versions of constructivist mathematics.

According to the constructivists, a mathematical object can only be said to exist if some procedure is laid down by which it can be constructed. If we apply this to infinite sequences, it looks as if we can only say that an infinite sequence exists if we can specify a rule for generating successive members of the sequence. Thus, for example, the decimal expansion of p can legitimately be said to exist because we can specify a rule for generating successive digits. But now consider whether, given this plausible constructive criterion, a random sequence can be said to exist. The answer would seem to be 'no'. Suppose using the constructive approach, we specify a rule for generating the sequence, then that rule could be used to give a successful gambling system, and hence the sequence would not be random. So we reach a strange result. Assuming classical mathematics, random sequences turn out to be more common than non-random sequences, since there exists a continuum infinity of them, and only a denumerable infinity of non-random sequences. Within some varieties of constructive mathematics, however, no random sequences exist. So do random sequences really exist or not? This is a difficult question which must be left for the reader to consider.

The relation between Von Mises' axioms and the Kolmogorov axioms*

Just as in the case of the subjective theory, we must now examine how Von Mises' axioms (as formulated above) relate to the Kolmogorov axioms, which are now standard among mathematicians. Let us therefore assume Von Mises' axioms, and see whether we can derive the Kolmogorov axioms. The first two axioms given in Chapter 4 are part of the Kolmogorov axioms and we will begin by showing that they can be derived from the axiom of convergence.

In stating the axioms of the previous chapter here, I will replace events E, F, ... by attributes A, B, ..., and the certain event by the attribute space. With these modifications we have:

Axiom 1

$0 \leq P(A) \leq 1$ for any A, and $P(\Omega) = 1$.

Assuming the axiom of convergence, we have $P(A) = \lim_{n \to \infty} m(A)/n$. Now $0 \leq m(A)/n \leq 1$. So, taking limits, $0 \leq P(A) \leq 1$. $m(\Omega)/n = n/n = 1$. So, taking limits, $P(\Omega) = 1$.

Axiom 2 (Addition Law)

If A, B are two exclusive attributes, then $P(A) + P(B) = P(A \vee B)$

If A, B are two exclusive attributes, then $m(A)/n + m(B)/n = m(A \vee B)/n$. So taking limits and using the axiom of convergence, we have $P(A) + P(B) = P(A \vee B)$, as required.

This demonstrates the Addition Law in the case of finite additivity. However, just as in the case of the subjective theory, we can raise the question of whether the Addition Law can be extended to countable additivity. In fact, countable additivity does not follow from Von Mises' axioms, as I will now show.[2] In order to investigate this problem, we have the initial difficulty that in any empirical collective the attribute space will be finite. It is not clear therefore how we can introduce infinite attribute spaces for which the question of countable additivity can be raised. Following Von Mises' general strategy, we need to look for a case in which the large finite can reasonably be approximated by the infinite. Let us consider a manufacturer of car engines who numbers each engine produced successively starting from 1. Suppose at a given moment we select at random a car which has an engine of this type and make a note of the engine number. Now at the time of the selection, some finite number N say of the engines will have been produced and fitted to cars. Thus the probability of selecting a number n for $1 \leq n \leq N$ is given by $1/N$. Suppose we select engine numbers successively in the same random fashion, since N is large, and indeed increasing, we could to a first approximation take N as infinite, i.e. regard the attribute space Ω as $\{1, 2, ..., n, ...\}$, and take $P(n) = 0$. Suppose we postulated countable additivity for the corresponding mathematical collective. Then we would have $P(\Omega) = P(\{1, 2, ..., n, ...\}) = P(1) + P(2) + ... + P(n) + ... = 0$. But, by Axiom 1, $P(\Omega) = 1$. This is a contradiction, which shows that countable additivity does not always hold for collectives which satisfy Von Mises' two axioms.

In his 1936 article on Von Mises' theory of probability, the issue of countable additivity is about the only one on which De Finetti agrees with Von Mises. De Finetti writes:

> And to end this, I still point out the agreement about a particular theorem: the extension of the theorem of total probabilities to denumerable classes, which is supported by many authors, on the contrary is not justified, not according to Von Mises' theory, nor to my viewpoint.
>
> (1936:364)

I have argued that countable additivity is in fact justified in De Finetti's theory, but that it is not justified in Von Mises' theory. Moreover, this seems to me a difficulty for Von Mises' theory, since it is in my view an advantage for any philosophical theory of probability to be able to justify the full mathematical apparatus currently used.

Von Mises was aware of this problem, and in his later work he attempted to resolve it by postulating countable additivity as a third axiom in addition to the axioms of convergence and randomness. (See Von Mises 1964a:12, where the axiom of countable additivity appears as equation (2).) This solves the problem from a mathematical point of view since the resulting theory is consistent. However, it undermines Von Mises' general philosophical justification of the axioms. According to Von Mises, each axiom should be the mathematical abstraction and idealisation of an empirical law. This account is plausible for the axioms of convergence and randomness, but does not apply at all to his extra axiom of countable additivity. In Chapter 7 I will show that this defect in Von Mises' theory can be overcome in the propensity theory of probability.

We must now consider:

Axiom 3 (Multiplication Law)

For any two attributes A, B, $P(A \,\&\, B) = P(A \mid B)\, P(B)$

As we saw in the previous chapter, Kolmogorov introduces conditional probabilities by a definition rather than an axiom. However, we argued that the use of an axiom was preferable. Formally speaking this does not make much difference. We have now to see if we can derive the above axiom in Von Mises' theory. To do so we must first devote a little question to the meaning of conditional probabilities within Von Mises' theory.

As we have seen (p. 97), the probability of any attribute A is, in Von Mises' theory, always conditional on some collective **C**, so that we should write $P(A \mid \mathbf{C})$. However, the conditional probability in Axiom 3 above is $P(A \mid B)$, which makes the probability of A conditional *not* on a collective *but* on an attribute B. Thus $P(A \mid B)$ has not so far been given a meaning, and we must do so before we can deal with Axiom 3. In fact $P(A \mid B)$ is defined as $P(A \mid B \,\&\, \mathbf{C})$, where B & **C** is a collective obtained from **C** as follows. We select from **C** those elements at which the attribute B occurs, and the resulting sequence is B & **C**. Of course we must next prove that B & **C** as just defined is indeed a collective, i.e. satisfies the axioms of convergence and randomness. We will in fact show this in the course of proving that Axiom 3 holds. There is, however, a preliminary point. Suppose B occurs only a finite number of times in the collective **C**. Then B & **C** will have only a finite number of members, and so, *a fortiori,* will not be a collective, which is an infinite sequence. Now if B occurs only a finite number of times in **C**, then $P(B \mid \mathbf{C}) = 0$. So, if we specify that $P(B \mid \mathbf{C}) \neq 0$, we eliminate this awkward case. In fact, however, the condition $P(B \mid \mathbf{C}) \neq 0$ is unnecessarily restrictive, because there might be a case in which B occurred infinitely often, and yet $P(B \mid \mathbf{C}) = 0$. However, to avoid mathematical complexities we will assume $P(B \mid \mathbf{C}) \neq 0$, though it should be noted that it is possible by complicating the mathematics to introduce probabilities conditional on attributes of zero probability within the frequency theory.

We can now proceed to our proof of Axiom 3, which will, at the same time, show that B & **C** is indeed a collective. Choose n arbitrarily and suppose that in the first n places of **C**, B occurs $n(B)$ times. Since $P(B \mid \mathbf{C}) \neq 0$ by assumption, $n(B) \to \infty$ as $n \to \infty$. Suppose in the first $n(B)$ places of B & **C**, A occurs $m(A)$ times, we have first to show that $\lim_{n(B) \to \infty} m(A)/n(B)$ exists. Now if A & B occurs $n(A \,\&\, B)$ times on the first n places of **C**, then $n(A \,\&\, B) = m(A)$. Hence

$$\lim_{n(\mathrm{B})\to\infty} \frac{m(\mathrm{A})}{n(\mathrm{B})} = \lim_{n\to\infty} \frac{n(\mathrm{A}\,\&\,\mathrm{B})}{n(\mathrm{B})} = \lim_{n\to\infty} \frac{n(\mathrm{A}\,\&\,\mathrm{B})/n}{n(B)/n}$$

So, by the axiom of convergence applied to **C**, we have

$$\lim_{n(\mathrm{B})\to\infty} \frac{m(\mathrm{A})}{n(\mathrm{B})} \quad \text{exists and equals} \quad \frac{\mathrm{P}(\mathrm{A}\,\&\,\mathrm{B})}{\mathrm{P}(\mathrm{B})}$$

To complete the proof, we have to show that this limit is unaltered by any recursive gambling system applied to B & **C**. Let g be such a system. Extend g to a recursive gambling system g′ for **C** as follows. Suppose B has so far appeared n - 1 times in **C**, then use the value of g(n) to either select or reject successive members of **C** until B occurs again. Then switch to g(n + 1), and so on. Let the collective selected by g′ be **C**′. The collective selected by g is B & **C**′. However, by the axiom of randomness applied to **C**, the limiting frequencies in **C**′ are the same as in **C**. Hence applying the first part of the proof, we have that the limiting frequencies of B & **C**′ exist and are the same as in B & **C**.

We have now shown that the Kolmogorov axioms follow in Von Mises' theory, if (a) we restrict ourselves to finite additivity, and (b) we limit Axiom 3 to the case where P(B) $\neq$ 0. One curious feature of the proof should be noted. The axiom of randomness was used only in the second half of the proof of Axiom 3, that is to check that B & **C** satisfied the axiom of randomness, and so was indeed a collective. If therefore we drop the axiom of randomness altogether and require only that collectives satisfy the axiom of convergence, then the Kolmogorov axioms (with the above restrictions) will still follow. To put the matter informally, all the Kolmogorov axioms seem to correspond to just the first of Von Mises' two axioms, and there is nothing in the Kolmogorov axioms corresponding to the axiom of randomness. This is certainly a strange situation. The axiom on which Von Mises laid such stress does not seem to appear at all in the standard mathematical axiomatisation. The reasons for this certainly need to be investigated further, but rather than doing so now, it will be convenient to postpone further discussion until Chapter 7, where it can be taken up in the context of the propensity theory.

6 The propensity theory

(I) General survey

The propensity theory of probability was introduced by Popper (1957b),[1] and subsequently expounded and developed by him in a series of papers and books (1959b, 1983, 1990). In *Logic of Scientific Discovery* (1934: Chapter VIII, 146–214), Popper advocated a version of the frequency theory. He continued to support an objective interpretation of probability, but subsequent reflection convinced him that the frequency theory was inadequate, and that therefore a new objective interpretation of probability was needed. This he sought to provide with his propensity theory. The main drawback of the frequency theory, according to Popper, was its failure to provide objective probabilities for single events. Yet he thought that these were needed for quantum mechanics.

Popper's suggestion of a propensity theory of probability has been taken up by quite a number of philosophers of science who have developed the idea in different ways. As a result there are now several different propensity theories. As Miller puts it:

> One of the principal challenges confronting any objectivist theory of scientific knowledge is to provide a satisfactory understanding of physical probabilities. The earliest ideas here, known collectively as the frequency interpretation of probability, have now been all but abandoned, and have been replaced by an equally diffuse set of proposals all calling themselves the propensity interpretation of probability.
>
> (1994:175)

In the case of the theories of probability so far considered (classical, logical, subjective and frequency), there has existed a more or less canonical version, and I have been able to concentrate on expounding this while noting some possible variations here and there. The situation with the propensity theory is very different. Here we have a 'diffuse set of proposals' which are currently being developed by different philosophers of science in different directions. This calls for a different expository technique.

The first step, which I will undertake in this chapter, is to give a general survey of propensity theories and indicate what problems they face. This survey will be far from complete, and I will confine myself to describing accounts of propensity due to Popper, Miller and Fetzer, as well as one of my own.[2] Although this is a limited selection from a larger menu, it should give a feeling for the varieties of propensity. Naturally, I will argue for my own propensity theory, but I hope to show as well some of the pros and cons of different accounts of propensity. The whole situation is quite intricate, and I will approach it historically by describing (pp. 114–18) Popper's first version of the propensity theory. Then in the section 'Can there be objective probabilities of single events?', I will consider whether this theory of Popper's really does solve the problem of providing objective probabilities for single events. My conclusion will be that it does not, and indeed that objective probabilities of single events may not be necessary at all. At this stage it may look as if I am abandoning the propensity theory altogether, but this is not the case. In the section 'Classification of propensity theories', I will suggest that we use the term 'propensity theory' not just for Popper's own theory, but for any theory which tries to develop an objective, but non-frequency, interpretation of probability. It seems to me that such an interpretation is needed for reasons which have nothing to do with the question of whether there are objective probabilities of single events. This analysis of propensity leads to a classification of propensity theories. In the section 'The propensity theories of Miller, the later Popper and Fetzer,' I consider the propensity theories of Miller and the later Popper, and of Fetzer. This leads to a further refinement of the classification introduced in 'Classification of Propensity Theories'. In the section 'Propensity and causality', I consider how the three kinds of propensity theory which have been introduced cope with one of the main problems confronting the whole approach. This problem concerns the relation between propensity and causality, and involves what is known as 'Humphreys' Paradox'. After this survey of some of the current propensity theories and the problems they face, I will turn in Chapter 7 to developing my own preferred version of the propensity theory, which, in the classification introduced in 'Classification of propensity theories' would be described as a 'long-run propensity theory'.

Popper's introduction of the propensity theory

The problem which gave rise to the propensity theory had already been considered by Popper in 1934. The question was whether it was possible to introduce probabilities for single events, or *singular probabilities* as Popper called them. Von Mises, assuming of course his frequency theory of probability, had denied that such probabilities could validly be introduced. The example he considered was the probability of death. We can certainly introduce the probability of death before the age of 41 in a sequence of say 40-year-old Englishmen. It is simply the limiting frequency of those in the sequence who die before age 41. But can we consider the probability of death before 41 for a particular 40-year-old Englishman (Mr Smith say)? Von Mises answered: 'no!':

> We can say nothing about the probability of death of an individual, even if we know his condition of life and health in detail. The phrase 'probability of

> death', when it refers to a single person has no meaning at all for us. This is one of the most important consequences of our definition of probability ...
> (1928:11)

Of course it is easy to introduce singular probabilities on the subjective theory. All Mr Smith's friends could, for example, take bets on his dying before age 41, and hence introduce subjective probabilities for this event. Clearly, however, this procedure would not satisfy an objectivist like Popper. The key question for him was whether it was possible to introduce objective probabilities for single events.

Popper in 1934 disagreed with Von Mises' denial of the possibility of objective singular probabilities, partly because he wanted such probabilities for his interpretation of quantum mechanics. Popper therefore considered a single event which was a member of one of Von Mises' collectives and made the simple suggestion that its singular probability might be taken as equal to its probability in the collective as a whole. Popper (1957b, 1959b) presented an objection, which he had himself invented, to this earlier view of his, and this led him to his new theory of probability.

Popper's argument is as follows. Begin by considering two dice: one regular, and the other biased so that the probability of getting a particular face (say the 5) is $^1/_4$. Now consider a sequence consisting almost entirely of throws of the biased die but with one or two throws of the regular die interspersed. Let us take one of these interspersed throws and ask what is the probability of getting a 5 on that throw. According to Popper's earlier suggestion this probability must be $^1/_4$ because the throw is part of a collective for which prob(5) = $^1/_4$. But this is an intuitive paradox, since it is surely much more reasonable to say that prob(5) = $^1/_6$ for any throw of the regular die.

One way out of the difficulty is to modify the concept of collective so that the sequence of throws of the biased die with some throws of the regular die interspersed is not a genuine collective. The problem then disappears. This is just what Popper did:

> All this means that the frequency theorist is forced to introduce a modification of his theory – apparently a very slight one. He will now say that an admissible sequence of events (a reference sequence, a 'collective') must always be a sequence of repeated experiments. Or more generally, he will say that admissible sequences must be either virtual or actual sequences which are *characterized by a set of generating conditions* – by a set of conditions whose repeated realisation produces the elements of the sequences.
> (1959b:34)

He then continued a few lines later: 'Yet, if we look more closely at this apparently slight modification, then we find that it amounts to a transition from the frequency interpretation to the propensity interpretation.' (Popper 1959b:34). In this interpretation, the generating conditions are considered as endowed with a propensity to produce the observed frequencies. As Popper put it: 'But this means that we have to visualise the conditions as endowed with a tendency or disposition, or propensity, to produce sequences whose frequencies are equal to the probabilities; which is precisely what the propensity interpretation asserts.' (1959b:35). There is an ambiguity in this formulation. Popper does not make it clear whether, when speaking of sequences, he means infinite sequences or long, but still finite, sequences. One piece of evidence in favour of the former interpretation is that Popper speaks of 'frequencies' being 'equal to the probabilities'. Now limiting frequencies in infinite sequences would be exactly equal to the probabilities, but frequencies in long finite sequences would only be *approximately* equal to the probabilities. There are however two pieces of evidence against the view that Popper had infinite sequences definitely in mind.

First of all in his exposition of the frequency theory earlier in the same paper Popper gives what is clearly an ambiguous formulation:

> From the point of view of the frequency interpretation, the probability of an *event of a certain kind* – such as obtaining a six with a particular die – can be *nothing but* the relative frequency of this kind of event in an extremely long (perhaps infinite) sequence of events.
>
> (1959b:29)

Second, the formulation of the propensity theory in the 1957b paper seems to favour the finite sequence interpretation. Popper says:

> ... since the probabilities turn out to depend upon the experimental arrangement, they may be looked upon as *properties of this arrangement. They characterize the disposition, or the propensity,* of the experimental arrangement to give rise to certain characteristic frequencies *when the experiment is often repeated.*
>
> (1957b:67)

Surely only finite sequences can be produced by experiments which are often repeated.

I do not intend to continue with a further exegesis of Popper. I introduced the point mainly to stress that throughout what follows I will adopt the *long but finite* sequences interpretation, and correspondingly regard the conditions as having a propensity to produce frequencies which are *approximately* equal to the probabilities. This is because my aim is to make the propensity theory more scientific and empirical, and it is obvious that infinite sequences of repetitions are not to be found in the empirical world. It may be objected to this interpretation that it is very difficult to say when a sequence of repetitions is long, or how close two numbers must become in order to be approximately equal. There is indeed a problem here, and I will discuss it in detail in the next chapter.

Popper's suggestion that probabilities should be related to the outcomes of sets of repeatable conditions (**S**) rather than collectives (**C**) had in fact already been made by Kolmogorov (1933). In the section which discusses the relations of his theory to experimental data (Kolmogorov 1933: Chapter 1, §2), he says in a footnote: 'In establishing the premises necessary for the applicability of the theory of probability to the world of actual events, the author has used, in large measure, the work of R. v. Mises.' (1933:3). In point of fact, however, Kolmogorov does not follow Von Mises in associating probabilities with collectives, but rather associates them with repeatable conditions, as the following quotation shows:

> There is assumed a complex of conditions, **S**, which allows of any number of repetitions.... If the variant of the events which has actually occurred upon realization of conditions **S** belongs to the set A (defined in any way), then we say that the event A has taken place.... Under certain conditions, ..., we may assume that to an event A which may or may not occur under conditions **S**, is assigned a real number P(A) ...
>
> (Kolmogorov 1933:3–4)

Kolmogorov did not, however, give any argument for his abandonment of Von Mises' concept of collective, and such an argument was supplied by Popper.

There is, however, rather more to Popper's notion of propensity than is involved in the change from collectives to conditions. The word 'propensity' suggests some kind of dispositional account, and this marks a difference from the frequency view. A useful way of looking into this matter will be to consider some earlier views of Peirce which were along the same lines.[3] These are contained in the following passage:

> I am, then, to define the meaning of the statement that the *probability,* that if a die be thrown from a dice box it will turn up a number divisible by three, is one-third. The statement means that the die has a certain "would-be"; and to say that the die has a "would-be" is to say that it has a property, quite analogous to any *habit* that a man might have. Only the "would-be" of the die is presumably as much simpler and more definite than the man's habit as the die's homogeneous composition and cubical shape is simpler than the nature of the man's nervous system and soul; and just as it would be necessary, in order to define a man's habit, to describe how it would lead him to behave and upon what sort of occasion – albeit this statement would be no means imply that the habit *consists* in that action – so to define the die's "would-be" it is necessary to say how it would lead the die to behave on an occasion that would bring out the full consequence of the "would-be"; and this statement will not of itself imply that the "would-be" of the die *consists* in such behavior.
>
> (Peirce 1910:79–80)

Peirce then goes on to describe 'an occasion that would being out the full consequence of the "would-be"'. Such an occasion is an infinite sequence of throws of the die and the relevant behaviour of the die is that the appropriate relative frequencies fluctuate round the value 1/3, gradually coming closer and closer to this value and eventually converging on it. Nothing is mentioned about ‘excluded gambling systems’.

Peirce is of course mistaken in speaking of the ‘would-be’ as a property of the die. Obviously it depends on the conditions under which the die is thrown, as is shown by the following two interesting examples of Popper’s. Suppose first we had a coin biased in favour of ‘heads’. If we tossed it in a lower gravitational field (say on the Moon), the bias would very likely have less effect and Prob(heads) would assume a lower value. This shows an analogy between probability and weight. We normally consider weight loosely as a property of a body whereas in reality it is a relational property of the body with respect to a gravitational field. Thus the weight of a body is different on the Moon whereas its mass (a genuine property of the body) is the same. For the second example we can use an ordinary coin, but this time, instead of letting it fall on a flat surface, say on a table top, we allow it to fall on a surface in which a large number of slots have been cut. We now no longer have two outcomes ‘heads’ and ‘tails’ but *three,* namely ‘heads’, ‘tails’ and ‘edge’; the third outcome being that the coin sticks in one of the slots. Further, because ‘edge’ will have a finite probability, the probability of ‘heads’ will be reduced. This example shows that not only do the probabilities of outcomes change with the manner of tossing but even that the exact nature of the outcomes can similarly vary.

Despite this error, Peirce has made what seems to me a valuable point in distinguishing between the probability of the die as a dispositional quantity, a ‘would-be’, on the one hand, and an occasion that would bring out the full consequence of the ‘would-be’ on the other. The importance of making this distinction is that it allows us to introduce probabilities as ‘would-be’s’ even on occasions where the full consequences of the ‘would-be’ are not manifested, where in effect we do not have a long sequence of repetitions. On the other hand, if we regard probabilities as ‘consisting in such behavior’ then it will only make sense to introduce probabilities on ‘occasions of full manifestation’, i.e. only for long sequences of repetitions. All this will become clearer if we now return to Von Mises and Popper.

It is a consequence of Von Mises’ position that probabilities ought only to be introduced in physical situations where we have an empirical collective, i.e. a long sequence of events whose outcomes obey the two familiar laws. If we adopt Popper’s propensity theory, however, it becomes perfectly legitimate to introduce probabilities on a set of conditions *even though these conditions are not repeated a large number of times.* We are allowed to postulate probabilities (and might even obtain testable consequences of such a postulation) when the relevant conditions are only repeated once or twice. Thus Popper’s propensity theory provides a valuable extension of the situations to which probability theory applies as compared to Von Mises’ frequency view. But does Popper’s propensity theory provide at the same time a solution to the problem of introducing objective probabilities for single events? We shall consider this question in the next section.

Can there be objective probabilities of single events?

What is perhaps the major difficulty in the way of introducing objective probabilities for single events was discussed by Ayer (1963:188–208), though the problem has an earlier history. The difficulty is this. Suppose we are trying to assign a probability to a particular event, then the probability will vary according to the set of conditions which the event is considered as instantiating – according, in effect, to how we describe the event. But then we are forced to consider the probabilities as attached to the conditions which describe the event rather than to the event itself.

To illustrate this, let us return to our example of the probability of a particular man aged 40 living to be 41. Intuitively the probability will vary depending on whether we regard the individual merely as a man or more particularly as an Englishman; for the life expectancy of Englishmen is higher than that of mankind as a whole. Similarly, the probability will alter depending on whether we regard the individual as an Englishman aged 40 or as an Englishman aged 40 who smokes two packets of cigarettes a day, and so on. This does seem to show that probabilities should be considered as dependent on the properties used to describe an event rather than as dependent on the event itself.

It is natural in the context of the propensity theory to consider the problem in terms of the conditions used to describe a particular event, but we could equally well look at the problem as being that of assigning the event to a reference class. Instead of asking whether we should regard Mr Smith as a man aged 40, as an Englishman aged 40 or as an Englishman aged 40 who smokes two packets of cigarettes a day, we could ask equivalently whether we should assign him to the reference class of all men aged 40, of all Englishmen aged 40 or of all Englishmen aged 40 who smoke two packets of cigarettes a day. The reference class formulation is more natural in the context of the frequency theory where the problem first appeared. Although we are discussing the propensity theory, we will continue to use the traditional terminology and refer to this fundamental problem as *the reference class problem.*

Howson and Urbach's (1989) reaction to the reference class problem is to argue that single case probabilities are subjective rather than objective. However, they also suggest that singular probabilities, though subjective, may be based on objective probabilities. Suppose, for example, that the only relevant information which Mr B has about Mr A is that Mr A is a 40-year-old Englishman. Suppose Mr B has a good estimate (p say) of the objective probability of 40-year-old Englishmen living to be 41. Then it would be reasonable for Mr B to put his subjective betting quotient on Mr A's living to be 41 equal to p, and thereby making his subjective probability objectively based. This does not, however, turn Mr B's subjective probability into an objective one, for consider Mr C, who knows that Mr A smokes two packets of cigarettes a day, and who also has a good estimate of the objective probability (q say) of 40-year-old Englishmen who smoke two packets of cigarettes a day living to be 41. Mr C will put his subjective probability on the same event (Mr A living to be 41) at a value q

different from Mr B's value *p*. Once again the probability depends on how the event is categorised rather than on the event itself. Howson and Urbach put the point as follows:

> ... single-case probabilities ... are not themselves objective. They are subjective probabilities, which considerations of consistency nevertheless dictate must be set equal to the objective probabilities just when all you know about the single case is that it is an instance of the relevant collective. Now this is in fact all that anybody ever wanted from a theory of single-case probabilities: they were to be equal to objective probabilities in just those conditions. The incoherent doctrine of objective single-case probabilities arose simply because people failed to mark the subtle distinction between the values of a probability being objectively based and the probability itself being an objective probability.
>
> (1989:228)

I am inclined to accept this criticism of Howson and Urbach, and so to adopt the following position. We can certainly introduce objective probabilities for events A which are the outcomes of some sets of repeatable conditions S. When, however, we want to introduce probabilities for single events, these probabilities, though sometimes objectively based, will nearly always fail to be fully objective because there will in most cases be a doubt about the way we should classify the event, and this will introduce a subjective element into the singular probability. I will now try to elaborate this position, and to discuss some further arguments which can be given in favour of objective singular probabilities. The first of these (the Ali–Holmes example) is due to Robert Northcott.

In 1980 Muhammad Ali, aged 38, fought Larry Holmes for the world heavyweight title. Because Muhammad Ali was a famous and popular figure, the majority of people accepted betting quotients in his favour which were too high, and so the punters made a lot of money by betting in favour of Larry Holmes, who won easily. Does this not indicate that there was an objective probability of Muhammad Ali winning which was much lower than most people thought?

I think that this argument does indeed establish something, but the conclusion is rather weaker than the existence of an objective singular probability. What it does show is that some subjective probabilities (betting quotients) may be preferable to others as a basis for action, but the existence of better subjective probabilities does not establish the existence of a single objective probability. The example is an instance of a general principle which may be roughly stated as follows. On the whole, it is better to use as the basis for action a subjective probability (betting quotient) based on more evidence rather than one based on less evidence. Thus in the Ali–Holmes example, the punters knew a great deal more than the general public did about the effects of age on a boxer's performance, on the relative form of Ali and Holmes, etc. So the subjective probability of Ali winning assigned by a punter was likely to have been a better basis for action than one assigned by an ignorant member of the public.

Let us next look at a particular instance of this general principle. Suppose a particular event E can be classified as an instance of a series of conditions **S**, **S′**, **S″**, ..., where the set of conditions **S** is a subset of **S′**, which is a subset of **S″**, and so on. Suppose further that statistical data enable us to obtain good estimates of the objective probability of E's occurring relative to **S**, **S′**, **S″**, ..., say p, p', p'', ... Then common sense suggests that it would, when considering the occurrence of E, be better to adopt as our probability p' rather than p, p'' rather than p', and so on. If instead of the conditions **S**, we consider the reference class of the set of instances of **S**, then the principle here could be called the *principle of the narrowest reference class.* It is regarded by Ayer as 'rational to accept'. He states it as follows:

> The rule is that in order to estimate the probability that a particular individual possesses a given property, we are to choose as our class of reference, among those to which the individual belongs, the narrowest class in which the property occurs with an extrapolable frequency.
>
> (1963:202)

Again we can illustrate this by our example of the probability of a particular 40-year-old man living to the age of 41. This individual can be put in the following reference classes: the class of 40-year-old men, the class of 40-year-old Englishmen, the class of 40-year-old Englishmen who smoke two packets of cigarettes a day. Now suppose we have good statistical data for all three classes, then the principle of the narrowest reference class suggests that we should base our probability of the particular individual living to be 41 on the frequency in the third of these three reference classes.

The principle of the narrowest reference class certainly seems to be a sound one, but there are some problems with it. First of all, there may not be a single narrowest reference class for which statistics are available.[4] Suppose Mr Smith in addition to smoking two packets of cigarettes a day plays football once a week. Let us suppose we have statistical data regarding death within a year for the class of 40-year-old Englishmen who smoke two packets of cigarettes a day and for the class of 40-year-old Englishmen who play football once a week, but *not* for the class of 40-year-old Englishmen who both smoke two packets of cigarettes a day and play football once a week. We thus have not one but two narrowest reference classes for which statistical data are available and the frequency estimates of the probability of Mr Smith living to the age of 41 on the bases of these two classes (p'', p''' say) may well be different.

Even if there is a single narrowest reference class, however, there may be, as Keynes pointed out, a danger in its uncritical use. Suppose we adopt the policy of taking as our probability for a single event the frequency ratio in the narrowest reference class to which that event belongs and for which good statistical data exist. Such a policy, according to Keynes, may well lead us astray because we may know things about the event which do not constitute statistical data in a reference class, but which, nonetheless, give us very good reasons for adjusting our probability. If we neglect such qualitative evidence and use only quantitative evidence, we may often be led to a probability which is a less satisfactory basis for action than might otherwise have been obtained. Keynes puts the point as follows:

> Bernoulli's second axiom, that in reckoning a probability we must take everything into account, is easily forgotten in these cases of statistical probabilities. The statistical result is so attractive in its definiteness that it leads us to forget the more vague though more important considerations which may be, in a given particular case, within our knowledge. To a stranger the probability that I shall send a letter to the post unstamped may be derived from the statistics of the Post Office; for me those figures would have but the slightest bearing upon the question.
>
> (1921:322)

Keynes obviously considered that he was either more likely than average to post an unstamped letter (perhaps through absent-mindedness or unconscious avarice) or less likely (through being very meticulous in his habits). He does not say which.

We can illustrate Keynes's point with our familiar example as follows. We are trying to assign a probability that our particular individual Mr Smith will live to be 41. Let us suppose that Mr Smith does not, after all, play football once a week and that there is a narrowest reference class for which we have good statistics, namely the class of 40-year-old Englishmen who smoke two packets of cigarettes per day. We accordingly estimate the probability of his living to be 41 as the frequency r say of those in this class who have lived to be 41. Suppose, however, that we learn that Mr Smith comes from a very numerous family who all smoke two packets of cigarettes per day, but none of whom has contracted lung cancer or any other smoking-related disease or indeed died before the age of 80. No statistical data are available concerning individuals who belong to such unusual families, but surely, in the light of this extra information, it would be reasonable to change our probability to a value somewhat higher than r.

The general procedure for assigning probabilities to single events then becomes something like the following. We first assign the event to the narrowest reference class for which reliable statistical data exist (if there is such a class) and calculate the relative frequency (r say) of the event's occurring in this class. We then consider any further information of a non-statistical character which is relevant to the event's occurring on this particular occasion, and adjust r either up or down in the light of this information to obtain our probability. If there happen to be several narrowest reference classes with relative frequencies r, r', r'', ... say, we then have to use the non-statistical information to choose a particular r-value as well as to adjust it. If there is no suitable reference class at all, we have to rely exclusively on the non-statistical information to decide on a subjective probability. Such a procedure is surely a reasonable and practical one, but it involves many subjective elements, and it is therefore unlikely to produce an objective singular probability in most cases. I will now give one further example of this procedure – *the Francesca argument.*

My wife is from Rome, and her sister has a daughter called Francesca. To explain how Francesca came to formulate the argument, some background on the social customs of Rome are needed. It seems that when a schoolchild reaches the age of 16 in Rome, it becomes necessary for emotional well-being and maintaining status with the peer group to own a motor scooter. Naturally, however, this causes great alarm to the parents (and even uncles), who are concerned about the possibility of a road accident. Francesca, when she became 16 was no exception to the general rule. So I had an argument with her on this subject. I pointed out that the frequency of 16-year-old Roman motor scooter riders who had accidents was quite high, and therefore that it might be better not to get a scooter. In her reply, Francesca accepted the truth of the statistics, and even added that two members of her school had already been taken to hospital in a coma as a result of motor scooter accidents. One girl had gone on her scooter without wearing a crash helmet to buy a pizza. She was returning balancing the pizza in one hand, and steering the scooter with the other, when the accident occurred. However, Francesca commented that this girl was extremely stupid, and that she (Francesca) would never do a thing like that. She would drive her scooter well and carefully, wear a crash helmet and take all the other recommended precautions, so that the probability of her having an accident was much lower than average. Although I was trying to support the opposite conclusion, it seemed to me that this argument of Francesca's could not be faulted. Indeed, it is a particular instance of Keynes's general principle. To one who knew her well, it did seem likely that she would drive well and carefully, and she would therefore be less likely to have an accident than the average 16-year-old Roman. The only criticism which might have been made is that accidents are sometimes the fault of the other party against whose errors even the very best and most careful driving offers no protection. Thus the reduction in the probability of an accident for a good and careful driver below the average level should perhaps not be too great.[5]

I will now consider one final argument in favour of objective probabilities for single events.[6] It might be conceded that it is difficult to assign such singular probabilities in cases like an individual dying before 41, or a 16-year-old Roman having an accident with her motor scooter. However, it could still be claimed that such singular probabilities are more plausible in cases like games of chance, or scientific experiments such as the quantum-mechanical two-slit experiment with electrons. Let us take games of chance first. Certainly in examples such as coin tossing or dice roiling, it does seem quite reasonable to say that on each toss or roll there is an objective singular probability equal to the objective probability in the sequence of tosses or rolls. Our earlier discussion shows why objective singular probabilities are more plausible here than in human cases involving individuals dying before 41 or having road accidents. In the human case there are many facts about the individual under consideration which do not take the form of statistical data relating to long sequences, but which seem relevant to assessing the probability. Perhaps there are strong indications that the individual in question has such a character as to make him or her a more careful driver than average, and so on. In the case of standard coin tossing, however, if fraud and malpractice are excluded, it is part of our background knowledge that additional facts about the toss do not influence the result. Thus, it does not matter whether the coin was heads uppermost or tails uppermost before it was tossed, whether it was allowed to fall on the table or on the floor, and so on. Our background knowledge therefore suggests that we should make the singular probabilities of each toss equal, and so equal to the objective probability in the sequence as a whole. Thus we could in this special case introduce objective singular probabilities, but there seems little point in doing so rather than accepting the Howson and Urbach analysis of a subjective probability based on an objective probability. After all, we would normally be interested in a particular toss (as opposed to a sequence of tosses) only if we wanted to gamble on the result, and thus the subjective probability analysis in terms of betting quotients seems quite appropriate.

Let us now consider scientific experiments. In my view there is a weaker case for introducing objective singular probabilities here than in the case of games of chance for the following reason. As just observed, it is characteristic of coin tossing and dice rolling that they can be carried out in a wide variety of ways without affecting the probability of getting a particular result. Indeed it is difficult, if not impossible, to toss a fair coin in such a way as to favour one side rather than the other.[7] With scientific experiments, however, the situation is quite different. It is often very difficult to perform the experiment correctly without extraneous factors disturbing the result. Great skill and care are needed to ensure that outside influences do not have an effect. Consider, for example, the quantum-mechanical two-slit experiment with electrons. Suppose that two scientists Mr A and Ms B are betting on where an electron will impinge in a particular repetition of the experiment. Mr A sets his probabilities equal to those calculated by the standard theory. Ms B, however, has noticed that there was a thunderstorm nearby, and knows from experience that the resulting electrical disturbances in the atmosphere often affect an experiment of this sort. She therefore adjusts her probabilities in the light of this factor. Once again it seems better to analyse the singular probabilities in a particular repetition of the experiment as subjective probabilities rather than as objective probabilities exactly equal to the objective probabilities in a sequence of repetitions of the experiment.

My general conclusion from the discussion of this section is as follows. It is reasonable in some cases to assign objective probabilities to events A which are the outcomes of sets **S** of repeatable conditions. Suppose Prob(A | **S**) = *p*. Popper claims that there is an objective singular probability *p* of A occurring on a particular instantiation of the conditions **S**. We have argued, however, that such a claim is justified, if at all, only in the case of simple games of chance such as coin tossing or dice rolling. In all other cases, and perhaps in the case of games of chance as well, it is more reasonable to analyse singular probabilities as subjective probabilities, which may, however, as Howson and Urbach have emphasised, be based at least partly on objective probabilities. As we have seen, Popper's propensity theory was developed in order to permit the introduction of objective singular probabilities. I have argued that it does not succeed in doing so, and it seems therefore as if I have thereby rejected the propensity theory. Certainly if propensity theory is used in a strict sense to describe Popper's precise views, then I do indeed reject the theory.[8] However, since Popper's introduction of the term 'propensity theory', it has come to have a wider significance and to mean roughly 'an objective, but non-frequency, theory'. In the next section I will examine this broader sense of the term 'propensity' and show how it leads to a classification of propensity theories. In the remainder of the chapter, the pros and cons of these various approaches to propensity will be considered.

Classification of propensity theories

A frequency theory of probability may be characterised as one in which probability is defined in terms of frequency either in the mathematical formalism or in an informal supplement designed to tie the theory in with experience. Now this indicates that frequency theories are based on an operationalist philosophy of science. Operationalism I take to be the view that the theoretical terms of a science should be defined in terms of observables. Frequency theories of probability are examples of the operationalist approach, because the theoretical term 'probability' is defined in terms of observable frequencies.

Now operationalism was very widely held in the 1920s but subsequently has been much criticised by philosophers of science. The alternative view which has come to prevail is that the theoretical terms of a natural science may often be introduced as undefined primitives and then connected to experience in a somewhat indirect fashion – not directly through an operational definition. If this more recent view is applied to probability theory, it ties in very nicely with the modern mathematical treatment of probability based on the Kolmogorov axioms. In Von Mises' mathematical treatment of probability, probability is explicitly defined in the mathematical formalism as limiting frequency. Kolmogorov abandons this approach, and in his mathematical development takes probability as a primitive undefined term which is characterised axiomatically. Admittedly Kolmogorov's approach is still compatible with a frequency theory, if we take probability as defined in terms of frequency in an informal supplement designed to connect the theory with experience. Indeed Kolmogorov (see 1933:§2, 3–5, including Footnote 4) himself seems to adopt a theory of this general character. Yet although Kolmogorov's mathematics is, in this sense, compatible with the frequency theory of probability, it seems to fit more naturally with recent non-operationalist philosophies of science, in which key theoretical concepts are often taken as undefined and are then connected only somewhat indirectly with observation.

These considerations, which, by the way, have nothing to do with the question of whether there are objective probabilities of single events, suggest that there is a need to develop an objective, but non-frequency, theory of probability. Such a theory would agree with Von Mises' view that probability theory is a mathematical science concerned with observable random phenomena. It would also agree with Von Mises' view that probability is an objective concept like mass in theoretical mechanics, or charge in electromagnetic theory. It would, however, differ from Von Mises' view that probability should be given an operational definition in terms of frequency. Probability would rather be introduced as a primitive undefined term characterised by a set of axioms[9] and then connected with observation in some manner more indirect than a definition in terms of frequency. My suggestion is that we should use Popper's term 'propensity theory' to describe any objective, but non-frequency, theory of probability having the general character just described.

Propensity theories in the above general sense can now be classified into (a) *long-run propensity theories* and (b) *single-case propensity theories.*[10] A long-run propensity theory is one in which propensities are associated with repeatable conditions, and are regarded as propensities to produce, in a long series of repetitions of these conditions, frequencies which are approximately equal to the probabilities. A single-case propensity theory is one in which propensities are regarded as propensities to produce a particular result on a specific occasion. As we have seen, Popper's original propensity theory was, in a sense, *both* long run *and* single case. His characterisation of propensities corresponds to our long-run propensities, and yet he wanted these propensities to apply to the single case as well. This position ran into difficulties connected with the reference class problem, and so there has been a tendency for the two halves of Popper's account to separate, producing two different types of propensity theory. My own preference is for a long-run propensity theory, and for dealing with the single case by subjective probabilities which may however be objectively based. But we will next examine more closely the other possibility of sticking to single-case propensities and modifying Popper's original account in other ways.

This analysis explains the fact, pointed out by Runde (1996), that Popper's later views on propensity, particularly Popper (1990), differ considerably from his earlier views. This later position is also developed by Miller (1994, 1996). It retains from the earlier Popper objective singular probabilities, but abandons the association of propensities with repeatable conditions. Instead propensities are associated with states of the universe. A single-case propensity theory was developed earlier by Fetzer (1981), but it differs significantly from the view of the later Popper and Miller. Instead of associating propensities with the complete state of the world at a given time, he associates them with a complete set of (nomically and/or causally) relevant conditions, which are subject to replication whether or not they are ever replicated. In the next section I will give a fuller account of these single-case propensity theories, and criticise them from the point of view of the long-run propensity theory advocated here.

The propensity theories of Miller, the later Popper and Fetzer

The main difference between the earlier and later Popper on propensity is that the earlier Popper associates propensities with repeatable conditions, whereas the later Popper says: '... propensities in physics are properties of *the whole physical situation* and sometimes of the particular way in which a situation changes.' (1990:17). One reason for this change may have been the desire to preserve objective probabilities for single events. If propensities are associated with repeatable conditions, then, as we argued in detail (pp. 119–25), it is difficult to carry them over to particular instances of these conditions. At all events Miller is determined to retain objective singular probabilities. He writes: ‘The principal virtue of the propensity interpretation in any of its variants is supposed to be that, unlike the frequency theory, it renders comprehensible single-case probabilities as well as probabilities in ensembles and in the long run.’ (Miller 1994:175), and again: ‘... the propensity interpretation ... is an objectivist interpretation where single-case probabilities are supreme’ (1994:177). Naturally I disagree with this point of view since I want to develop a version of the propensity theory in which objective single-case probabilities are abandoned.

In his earlier period, Popper wrote in a passage already quoted: ‘But this means that we have to visualise the conditions as endowed with a tendency, or disposition, or propensity, to produce sequences whose frequencies are equal to the probabilities; which is precisely what the propensity interpretation asserts.’ (Popper 1959b:35). As already explained, I would replace ‘equal’ here with ‘approximately equal’, but otherwise accept this passage as part of my own version of the propensity theory. Miller, however, criticises the view that propensities are propensities to produce frequencies. He regards them instead as propensities to realise particular outcomes. As he says: ‘In the propensity interpretation, the probability of an outcome is not a measure of any frequency, but (as will be explained) a measure of the inclination of the current state of affairs to realize that outcome.’ (Miller 1994:182). In a significant passage, Miller relates these changes from the position of the earlier Popper with the need to solve the problem of singular probabilities. As he says:

> It is to be regretted, therefore, that ... we find remarks [in the earlier Popper] that ... depict propensities as “tendencies to produce relative frequencies on repetition of similar conditions or circumstances” ... Propensities are not located in physical things, and not in local situations either. Strictly, every propensity (absolute or conditional) must be referred to the complete situation of the universe (or the light-cone) at the time. Propensities depend on the situation today, not on other situations, however similar. Only in this way do we attain the specificity required to resolve the problem of the single case.
> (Miller 1994:185–6)

That concludes my account of this version of the propensity theory, and I will now proceed to criticise it.

The main problem with the 1990s views on propensity of Popper and Miller is that they appear to change the propensity theory from a scientific to a metaphysical theory. If propensities are ascribed to a set of repeatable conditions, then by repeating the conditions we can obtain frequencies which can be used to test the propensity assignment. If, on the other hand, propensities are ascribed to the ‘complete situation of the universe ... at the time’, it is difficult, in view of the unique and unrepeatable character of this situation, to see how such propensity assignments could be tested. Miller seems to agree with this conclusion since he writes: ‘The propensity interpretation of probability is inescapably metaphysical, not only because many propensities are postulated that are not open to empirical evaluation' (1996:139). Popper too writes in similar vein:

> But in many kinds of events ... the propensities cannot be measured because the relevant situation changes and cannot be repeated. This would hold, for example, for the different propensities of some of our evolutionary predecessors to give rise to chimpanzees and to ourselves. Propensities of this kind are, of course, not measurable, since the situation cannot be repeated. It is unique. Nevertheless, there is nothing to prevent us from supposing such propensities exist, and from estimating them speculatively.
>
> (1990:17)

Of course we can indeed estimate the propensities speculatively, but if these speculations cannot be tested against data, they are metaphysical in character.

Now there is nothing wrong with developing a metaphysical theory of propensities, and such a theory may be relevant to the discussion of old metaphysical questions such as the problem of determinism. However, my own aim is to develop a propensity theory of probability which can be used to provide an interpretation of the probabilities which appear in such natural sciences as physics and biology. For a theory of this kind, probability assignments should be testable by empirical data, and this makes it desirable that they should be associated with repeatable conditions.

Fetzer's single-case propensity theory differs from that of Miller and the later Popper in that he does not associate propensities with the complete state of the universe. As Fetzer says:

> ... it should not be thought that propensities for outcomes ... depend, in general, upon the complete state of the world at a time rather than upon a complete set of (nomically and/or causally) relevant conditions. . which happens to be instantiated in that world at that time.
>
> (1982:195)

This seems to me a step in the right direction relative to Miller and the later Popper, but some doubts remain in my mind about how scientifically testable such propensities can be. If propensities are associated with a set of repeatable conditions as in the long-run propensity view, then it is always in principle possible to test a conjectured propensity value by repeating the conditions. If, as Fetzer suggests, we ascribe propensities to a complete set of (nomically and/or causally) relevant conditions, then in order to test a conjectured propensity value we must make a conjecture about the complete list of the conditions which are relevant. This necessary conjecture might often be difficult to formulate and hard to test, thereby rendering the corresponding propensities metaphysical rather than scientific. Once again then I have a doubt whether single-case propensities give an appropriate analysis of the objective probabilities which appear in the natural sciences.

On the other hand, it should be said in favour of Fetzer's view that, if the problem of finding the complete set of relevant conditions could be solved, his theory would provide an elegant and unified account. If we can define single-case propensities relative to some complete set $\mathbf{S}_C$ say of relevant conditions, we can then extend this to long-run sequences (whether finite or infinite) produced by repetitions of $\mathbf{S}_C$. My own account relates propensities to long-run repetitions of sets of conditions **S** which may not be complete. This makes propensity assignments easily testable, but means, for the reasons explained earlier (pp. 119– 25), that they cannot in general be extended to the single case where subjective probabilities are needed. Thus Fetzer's theory, if it could be made to work, would lead to a unified monistic account, whereas the alternative long-run propensity approach leads necessarily to a more complicated dualism.

We can now extend our classification of propensity theories by subdividing single-case propensity theories into (a) *state of the universe,* where the propensity depends on the complete state of the universe at a given time, and (b) *relevant conditions,* where the propensity depends on a complete set of relevant conditions. Miller and the later Popper opt for (a) and Fetzer for (b). If we add the long-run propensity theory here advocated, we have three different propensity theories. In the next section I will test out these three theories by seeing how well they deal with another major problem connected with propensity. This is the problem of relating propensity to causality, a problem which leads to what is known as Humphreys' paradox.

Propensity and causality: Humphreys' paradox

In his 1990 *A World of Propensities,* Popper made the interesting suggestion that propensity might be a generalisation of the notion of cause. As Popper puts it: 'Causation is just a special case of propensity: the case of a propensity equal to 1' (1990:20). Thus, to take a simple example, a large dose of cyanide will definitely cause death. A suitably small dose of cyanide might only give rise to a propensity of 0.6 of dying. So propensity appears to be a kind of weakened form of causality. A basic difficulty with the idea that propensities are generalisations of causes is the following. Causes have a definite direction in time. So if A causes B and A occurs before B, then B does not cause A. Apart from a few speculations in theoretical physics, it is universally conceded that causes do not operate backwards in time. The situation is very different with probabilities. For events A, B, we usually have that if $P(A \mid B)$ is defined, then so is $P(B \mid A)$. Probabilities have a symmetry where causes are asymmetrical. It thus seems that propensity cannot after all be a generalisation of cause.

This problem was first noticed by Humphreys, and first published by Salmon, who gave it a memorable formulation that is worth quoting:

> As Paul W. Humphreys has pointed out in a private communication, there is an important limitation upon identifying propensities with probabilities, for we do not seem to have propensities to match up with "inverse" probabilities.

> Given suitable "direct" probabilities we can, for example, use Bayes's theorem to compute the probability of a particular cause of death. Suppose we are given a set of probabilities from which we can deduce that the probability that a certain person died as a result of being shot through the head is $^3/_4$. It would be strange, under these circumstances, to say that this corpse has a propensity (tendency?) of $^3/_4$ to have had its skull perforated by a bullet. Propensity can, I think, be a useful causal concept in the context of a probabilistic theory of causation, but if it is used in that way, it seems to inherit the temporal asymmetry of the causal relation.
>
> (1979:213–14)

The problem was named 'Humphreys' paradox' by Fetzer (1981), and it has given rise to much interesting discussion. Humphreys (1985) himself gives a statement of the paradox, which is critically discussed by McCurdy (1996). There are also important contributions by Fetzer and Miller, which will be discussed later on. My aim, however, is not to give a complete review of the literature on the subject, but rather to carry out the following more limited strategy. I will begin by giving what seem to me the two simplest illustrations of the paradox. I will then examine how these cases can be dealt with by the three propensity theories which have been introduced earlier.

The first illustration comes from Milne (1986). Let us consider rolling a standard die, let A = 6 and B = even. Then, according to standard probability theory, $P(B \mid A) = 1$ and $P(A \mid B) = {^1/_3}$. $P(B \mid A)$ raises no problems. If the result of a particular roll of the die is 6, then that result must be even. B is completely determined by A, which corresponds satisfactorily to a propensity of 1. But now Milne (1986: 130) raises the question of how $P(A \mid B)$ is to be interpreted by a single-case propensity theory. There is indeed a problem here, for we cannot interpret it as saying that the occurrence of the outcome B partially causes with weight $^1/_3$ outcome A to appear. In fact if outcome B has occurred, then the actual result must have been 2, 4 or 6. In the first two cases it has been determined that 6 will not occur on that roll, while in the third case it has been determined that 6 will definitely occur. In neither case does it make any sense to say that A is partially determined to degree $^1/_3$ by the occurrence of B.

Milne's example concerns two events A and B which occur simultaneously. Yet characteristically causes come before their effects. We need, therefore, to consider $P(A \mid B)$ when A occurs at a different time from B. Cases of this sort are a little more complicated than Milne's simple die-rolling example. I have chosen what seems to me the simplest and most elegant. This is the frisbee example (Earman and Salmon 1992:70). It is essentially the same as an example involving electric light bulbs given previously by Salmon (1979:214).

Let us suppose then that there are two machines producing frisbees. Machine 1 produces 800 per day with 1 per cent defective. Machine 2, an older and less efficient machine, produces 200 per day with 2 per cent defective. Let us suppose that, at the end of each day, a frisbee is selected at random from the 1,000 produced by the two machines. Let D = the selected frisbee is defective. Let M = it was produced by machine 1, and N = it was produced by machine 2. Let us consider the two conditional probabilities P(D | M) and P(M | D). P(D | M) = 0.01, while P(M | D) can be calculated using Bayes's theorem as follows.

$$\begin{aligned} P(M \mid D) &= \frac{P(D \mid M)P(M)}{P(D \mid M)P(M) + P(D \mid N)P(N)} \\ &= \frac{0.01 \times 0.8}{0.01 \times 0.8 + 0.02 \times 0.2} \\ &= \frac{8}{12} = \tfrac{2}{3} \end{aligned}$$

As far as the standard operations of the calculus of probabilities are concerned, there is nothing problematic about these two conditional probabilities. But how are they to be interpreted in terms of single-case propensities, if D, M and N refer to a particular day?

Of course as regards P(D | M) there is no problem. This is just the propensity for machine 1 to produce a defective frisbee. But what of P(M | D)? This is the propensity for the actual defective frisbee drawn at the end of a particular day to have been produced by machine 1. If we think of propensities as partial causes, this becomes the following. The drawing of a defective frisbee in the evening is a partial cause of weight $^2/_3$ of its having been produced by machine 1 earlier in the day. Such a concept seems to be nonsense, because by the time the frisbee was selected, it would either definitely have been produced by machine 1 or definitely not have been produced by that machine. We can make the point more vivid by supposing that machine 1 produces blue frisbees and machine 2 red frisbees. If the defective frisbee drawn at the end of the day was blue, it would definitely have been produced by machine 1, and it is not clear what would be the sense of saying that it had a propensity $^2/_3$ to have been produced by machine 1. Obviously examples of this sort pose a problem for the propensity view of probability. Let us next examine how the three propensity theories so far discussed cope with the difficulty.

I will begin with the long-run propensity theory. Here propensities are associated with sets of repeatable conditions. Let **S** be such a set. Let the specific outcomes of **S** be members of a class Ω. Then propensities are assigned to events A, B, ... which are taken to be subsets of Ω. So, for example, P(A | **S**) = *p* means that there is a propensity if *S* were to be repeated a large number of times for A to appear with a relative frequency approximately equal to *p*. This view of propensity does not deal in any way with single repetitions of **S**. These are handled using subjective probabilities.

It is obvious form the above brief summary of the long-run propensity theory that basic propensities are conditional, and we have the form P(A | **S**) where **S** is a set of repeatable conditions. Note that here we cannot reverse the order because P(**S** | A) does not make sense. Now often for brevity the reference to **S** is not made explicit, and we abbreviate P(A | **S**) to P(A). Probabilities like P(A) are often called *absolute probabilities,* but really since P(A) is an abbreviation for P(A | **S**) it would be more accurate to refer to then as *fundamental conditional probabilities,* or *conditional probabilities in the fundamental sense.* Such fundamental conditional probabilities can be contrasted with conditional probabilities of the form P(A | B), where B is not a set of repeatable conditions but an event. It is these conditional probabilities which can be reversed to produce P(B | A). Let us call such conditional probabilities *event-conditional probabilities.* This gives rise to two questions. What do such conditional probabilities mean in the given interpretation? And how do they relate to fundamental conditional probabilities? I will now try to answer these questions.

My suggestion is that, just as P(A) should be considered as an abbreviation of P(A | **S**), so P(A | B) should be considered as an abbreviation of P(A | B & **S**) where B & **S** stands for a new set of repeatable conditions defined as follows. We repeat **S** just as before, but only note the result if it is a member of B. Results which do not lie in B are simply ignored. To say that P(A | B & **S**) = q means that there is a propensity if this new set of conditions B & **S** is repeated a large number of times for A to appear with frequency approximately equal to q. My next claim is that, with this interpretation of event-conditional probabilities, all the conditional probabilities in both the Milne example and the frisbee example make perfect sense and do not raise any problems.

Let us start with Milne's example in which the problem lay in interpreting P(A | B) = $^1/_3$, where A = the result of a roll of the die was 6, and B = the result of a roll of the die was even. The meaning of this probability on our long-run propensity view is the following. Suppose we roll the die a large number of times but ignore all odd results. There is a propensity, under these conditions, for 6 to appear with a frequency approximately equal to $^1/_3$. This is both true and straightforward. Note that Milne's difficulties disappear because we are considering the long run and not a single roll of the die.

The frisbee example is no more problematic than Milne's on this long-run propensity interpretation. Let **S** be the set of repeatable conditions specifying that the two machines produce their daily output of frisbees, and that, in the evening, one of these frisbees is selected at random and examined to see if it is defective. **S** can obviously be repeated each day. P(M | D) is now interpreted as an abbreviation for P(M | D & **S**). The statement P(M | D & **S**) = $^2/_3$ means the following. Suppose we repeat S each day, but only note those days in which the frisbee selected is defective, then, relative to these conditions, there is a propensity that if they are instantiated a large number of times M will occur, i.e. the frisbee will have been produced by machine 1, with a frequency approximately equal to $^2/_3$. Note that once again the difficulties disappear because we are considering the long run rather than a single case. In a specific instance it did not make sense to speak of the propensity of the selected frisbee having been produced by machine 1 as equal to $^2/_3$. If the selected frisbee was blue, it would have been produced by machine 1. If red, by machine 2. In either case the situation would have already been determined so that a propensity of $^2/_3$ would not make sense. If propensities are propensities to produce long-run frequencies, then a propensity of $^2/_3$ makes perfect sense, even though we know that in each individual case the result has been definitely determined as either M or N by the time the selected frisbee is examined and found to be defective.

So far, I have not mentioned the connection between causality and propensity, and the long-run propensity theory does seem to sever this connection. But is it wrong to do so? It is after all standard in discussions of causality to distinguish between causes and correlations. My barometer's falling sharply is very well correlated with rain occurring soon, but no one supposes that my barometer's falling sharply is the cause of the rain. Now correlation is a probabilistic notion. So perhaps it is indeed correct that causes are different from probabilities. On the long-run propensity interpretation, there is a high propensity of rain occurring soon, given that my barometer has fallen sharply, but this propensity is not causal in character. This concludes my discussion of Humphreys' paradox in the context of the long-run propensity theory. I will next examine how the paradox might be resolved within the two single-case propensity theories.

Let us begin with the state of the universe version of the single-case propensity theory developed by Miller and the later Popper. In this approach the probability of a particular event A is considered as conditional on an earlier state of the universe U_t say. $P(A \mid U_t) = p$ means that the state of the universe U_t has a propensity p to produce the event A. Here propensity is definitely thought of as a generalised cause. It will be sufficient to examine how this account applies in the frisbee case, since Milne's example does not raise any additional problems.

To deal adequately with the frisbee example within this propensity theory, it is important to add time subscripts to all the events involved. Let U_t be the state of the universe at the beginning of a particular day. Let D_v be the event that the frisbee drawn in the evening at time v was defective. Let M_u be the event that this defective frisbee was produced by machine 1 at time u during the day. Obviously we have $t < u < v$. The problem now is how to interpret the two conditional probabilities $P(D_v \mid M_u)$ and $P(M_u \mid D_v)$. As in the previous case, all probabilities are conditional. If we write P(A), this can only be an abbreviation for $P(A \mid U_t)$, where U_t is a state of the universe. So the fundamental conditional probabilities in this theory have the form $P(A \mid U_t)$. In the case of $P(D_v \mid M_u)$ and $P(M_u \mid D_v)$, however, neither M_u nor D_v are states of the universe but particular events. We are dealing with event-conditional probabilities, and once again we have to examine what sense can be made of these in the present theory.

Let us first try interpreting $P(D_v \mid M_u)$ as $P(D_v \mid M_u \,\&\, U_t)$ by analogy with what we did last time. The problem is that $M_u \,\&\, U_t$ is not a state of the universe at a particular time. Suppose we wanted to turn it into a state of the universe at the later time u (U_u say), then we would need to specify not just one event occurring at u such as M_u but all the other events occurring at u. Worse still, M_u might remain the same while these other events were different producing different values for U_u, and hence different values for $P(D_v \mid U_u)$. For example, in one U_u there might be severe oscillations in the supply of electricity between t and u increasing the proportion of defective frisbees produced by the two machines. In another U_u there might be no such disruption. The values of $P(D_v \mid U_u)$ would be different in the two cases.

There thus seems no hope of interpreting $P(D_v \mid M_u)$ along the lines we used in the previous case. The situation is even worse for $P(M_u \mid D_v)$, since if we extended D_v & U_t to U_v, then M_u would occur at a time earlier than v, which disallows $P(M_u \mid U_v)$. As far as I can see, there is only one way of introducing event-conditional probabilities in this theory, and that is by defining them formally thus.

$$P(A \mid B; U_t) = \text{def} \frac{P(A \,\&\, B \mid U_t)}{P(B \mid U_t)} \quad \text{for } P(B \mid U_t) \neq 0$$

But these formal event-conditional probabilities do not share the important properties of the fundamental conditional probabilities [$P(A \mid U_t)$] which underlie this version of the propensity theory. In particular, '$P(A \mid B) = p$' does not imply that there is any link of a causal character between B and A. Thus Humphreys' paradox on this approach is again solved by denying that event-conditional probabilities involve any kind of causal link, though, in contrast to the previous theory, it is maintained at the same time that conditional probabilities in the fundamental sense do involve a sort of causal link.

What I have just given is my own analysis of the situation, but it seems to agree quite well with what Miller says in the following passage:

> ... if *a* is my survival one year from today, and *c* is my taking up parachuting tomorrow, ... the causal influence that is measured by $p(a \mid c)$ is an influence from today to a day one year hence ... It is not an influence from the time recorded in *c* to the time recorded in *a*.... What about the inverse conditional probability $p(c \mid a)$? This comes out as the propensity for today's world to develop into a world in which I take up parachuting tomorrow, given that it – today's world – will by the end of the year have developed into one of the worlds in which I am still alive.... The causal pressure is from today to tomorrow, not from the remote future to tomorrow.
>
> (1994:189)

I have some doubts about the ordinary language equivalent given here for $p(c \mid a)$, but the point which agrees with my own analysis is that in $p(c \mid a)$ it is denied that there is causal pressure from *a* to *c,* and in $p(a \mid c)$ it is denied that there is influence from *c* to *a.* In effect it is denied that event-conditional probabilities involve any causal-type influences between the events. The influence is claimed to be from today to a day one year hence, or from today to tomorrow. In other words the causal type pressure runs from the state of the universe today to events lying in the future. It does not connect the future events which are involved in the event-conditional probability.

Let us finally examine Humphreys' paradox within the context of the relevant conditions single-case propensity theory due to Fetzer. As in the previous case we will focus on the frisbee example using the time subscripts already introduced. As in the previous two cases, we can distinguish between conditional probabilities in the fundamental sense and event-conditional probabilities. Within this theory, fundamental conditional probabilities have the form $P(A \mid R_t)$, where R_t is a complete set of (nomically and/or causally) relevant conditions instantiated in the world at time *t*. R_t consists of all those conditions which are relevant to the occurrence of the events under consideration, but it does not amount to a complete state of the universe at *t*. This is where Fetzer's version of the single-case propensity theory differs from that of Miller and the later Popper. $P(A \mid R_t) = p$ means that there is a propensity of degree *p* for the conditions R_t to produce the event A at some time later than *t*. Here, as in the previous case, propensity is thought of as a generalised cause.

Let us as before raise the question of how the two event-conditional probabilities $P(D_v \mid M_u)$ and $P(M_u \mid D_v)$ are to be interpreted in this theory. In fact, $P(D_v \mid M_u)$ can be interpreted quite straightforwardly as $P(D_v \mid M_u \,\&\, R_t)$. The set of relevant conditions at *t* is $R_{t.}$ At the later time *u*, however, the frisbee which will eventually be selected is produced by machine 1. This is the event M_u. The occurrence of this event is part of the set of relevant conditions for D_v at *u*. Thus we have $P(D_v \mid M_u) = P(D_v \mid M_u \,\&\, R_t) = P(A \mid R_u)$. So this event-conditional probability can be interpreted as a fundamental conditional probability at a different time, and therefore has the causal influence of a fundamental conditional probability. The inverse event-conditional probability $P(M_u \mid D_v)$ cannot however be interpreted in this way. If we tried to extend $D_v \,\&\, R_t$ to R_v, then $P(M_u \mid R_v)$ would no longer make sense as a generalised cause propensity since *v* is later than *u* and the direction of causality is wrong. Of course such event-conditional probabilities could still be introduced in a formal sense by a definition analogous to the one given earlier.

Humphreys' paradox is thus resolved within Fetzer's theory by saying that some, but not all, event-conditional probabilities are propensities (in the sense of generalised causes). This is how Fetzer puts it: '... by virtue of their "causal directedness", propensities cannot be properly formalized either as "absolute" or as "conditional" probabilities satisfying inverse as well as direct probability relations.' (1982:195), and again: '... that propensities are not probabilities (in the sense of satisfying standard axioms, such as Bayes's theorem) by virtue of their causal directedness was not generally recognized before the publication of Humphreys (1985).' (Fetzer 1991:297–8).

On Fetzer's account then, propensities do not satisfy the standard Kolmogorov axioms. Working with Nute, however, Fetzer developed an alternative set of axioms for propensities. This system which he calls 'a probabilistic causal calculus' is presented in his 1981 book (pp. 59–67). It has the feature that: '...*p* may bring about *q* with the strength *n* (where *p* occurs prior to or simultaneous with *q*), whether or not *q* brings about *p* with *any* strength *m* ...' (Fetzer 1981:284).

Fetzer's position might seem to be that propensities are not probabilities, but he objects to this formulation on the grounds that the Fetzer–Nute probabilistic causal calculus has many axioms which are definitely probabilistic in character. It might therefore be more accurate to describe the Fetzer–Nute calculus as a non-standard probability theory. As Fetzer himself says:

> Perhaps this means that the propensity construction has to be classified as a *non-standard* conception of probability, which does not preclude its importance *even as an interpretation of probability*! Non-Euclidean geometry first emerged as a *non-standard* conception of geometry, but its significance is none the less for that. Perhaps, therefore, the propensity construction of probability stands to standard accounts of probability just as non-Euclidean constructions of geometry stood to standard accounts of geometry before the advent of special and of general relativity.
>
> (1981:285)

Since 'non-standard' has connotations of 'non-standard analysis' it might be better to speak of the probabilistic causal calculus as a non-Kolmogorovian probability theory by analogy with non-Euclidean geometry.

The Fetzer–Nute suggestion of a non-Kolmogorovian probability theory is a bold and revolutionary one, but its revolutionary character will naturally create problems in its achieving general acceptance. There is an enormous body of theorems based on the Kolmogorov axioms. The mathematical community is unlikely to give up this formidable structure and substitute another for it unless there are very considerable gains in so doing. This is one reason for preferring a propensity theory (such as the long-run propensity theory) which retains the standard Kolmogorov axioms.

That concludes my general survey of propensity theories. Although each of the various approaches has some attractive features, my own preference is for the long-run propensity theory, and I will accordingly develop a particular version of this type of propensity theory in detail in the next chapter.

7 The propensity theory

(II) Development of a particular version

After the overview of propensity theories in the previous chapter, I will now try to develop a particular propensity theory in detail. This theory belongs to the type which was classified as *long-run propensity.* It is closer therefore to Popper's early ideas on propensity than to his later views. It is also the closest of the various propensity theories to Von Mises' frequency theory. The two views have in common the following points: (1) probability theory is a mathematical science like mechanics or electromagnetic theory; (2) this science deals with random phenomena to be found in the material world; (3) its axioms are confirmed empirically; and (4) probabilities exist objectively in the material world, like masses or charges, and have definite, though perhaps unknown, values.

Despite these similarities, there are of course differences between the two theories. First of all probabilities, being propensities, are associated with repeatable conditions rather than collectives. A second point concerns the relationship between the axioms of probability theory and Von Mises' two empirical laws. Von Mises regarded the axioms as obtained from these laws by a process of abstraction or idealisation. In the present version of the propensity theory, however, the axioms are regarded as explaining, and rendering more precise, the empirical laws. This conception, as we shall see, resolves some of the problems which we noted in Von Mises' account. It is also related to the third very crucial difference. The frequency theory gives a definition of the theoretical concept (probability) in terms of an observable quantity (frequency). It is thus based on an operationalist philosophy of science. The following version of the propensity theory is, on the contrary, based on a non-operationalist approach, according to which theoretical concepts are not defined in terms of observables. They are introduced as undefined notions which may be characterised axiomatically and are then connected to observables in a somewhat indirect fashion. In order to develop my version of the propensity theory, it becomes necessary to elaborate and define a non-operationalist theory of conceptual innovation in the natural sciences, and this will be done in the next section ('Criticisms of operationalism'). Before embarking on this, however, I would like to make one further point about operationalism.

As was pointed out in Chapter 4 (p. 58), the subjective theory of probability is based on operationalism. I am now, however, proposing to develop a version of the propensity theory which is based on a non-operationalist account which explicitly rejects operationalism.

At first sight this might not seem to be a serious problem. After all, the subjective theory and the propensity theory are fundamentally different. The subjective theory regards probabilities as degrees of belief and is hence epistemological, whereas the propensity theory sees probabilities as features of the material world, and so is objective. Granted these profound differences, why is there a problem if these two theories have radically different foundations? Indeed this is what we might expect.

Certainly for someone who supported just one of the two theories and rejected the other, there would be no problem. I have, however, argued in Chapter 6 that subjective probabilities are needed, in addition to propensities, to deal with the single case. This view will be elaborated in Chapter 8, in which I will argue for a pluralist view of probability, according to which there are different interpretations of probability which are valid and applicable in different circumstances. It will be further argued that these valid interpretations include both the subjective theory and the propensity theory of the present chapter. Since the former theory is based on operationalism, and the latter on a rejection of operationalism, it would seem that my pluralism involves both an acceptance and rejection of operationalism. This is indeed a serious problem, but this is not the appropriate place in which to try to tackle it. I have therefore to ask the reader to keep this difficulty at the back of his or her mind for the moment. I will return to the problem and attempt to resolve it in the last chapter (pp. 200–5).

Criticisms of operationalism:a non-operationalist theory of conceptual innovation in the natural sciences[1]

Since one of my main criticisms of the frequency theory is that it is based on what I regard as an inadequate operationalist philosophy of the natural sciences, I will begin my attempt to develop an objective, but non-frequency, theory of probability by criticising operationalism and putting forward a different theory of conceptual innovation in the natural sciences. As we have seen, Von Mises based his operationalist account of probability on Mach's earlier operationalist definition of Newtonian mass. In a similar fashion I will base my non-operationalist, propensity account of probability on the non-operationalist theory of conceptual innovation in the natural sciences presented in this section. To make the parallel closer, I will illustrate this non-operationalist theory with the example of Newtonian mass.

According to operationalism, the theoretical concepts of natural science should be defined in terms of observable concepts. We can put the claim in a more dynamic way by saying that every new concept introduced into natural science must be given an operational definition in terms of observational or experimental procedures. Thus, for example, the concept of length could be introduced by specifying a measuring procedure with rigid metre rods. Let us now look at some of the difficulties which such an account faces.

The first problem is that a single operational definition does not in general suffice for all the applications of a concept. Thus with our example of length, the rigid metre rod procedure may be adequate for lengths ranging from a few centimetres to a few hundred metres, but what about the distance between the Earth and the Sun? Or again, what about the diameter of an electron moving with a velocity near that of light? It would seem that we must introduce a sequence of operational definitions. Of course our different operational definitions must agree where they overlap, but there is another complication. Let us take the first simple extension of the concept of length. Suppose we wish to measure large terrestrial distances of the order of several kilometres. We have to supplement our use of metre sticks with theodolites. Now to use these instruments we have to make certain *theoretical* assumptions. For example, we must assume that light rays move in straight lines and that space is Euclidean. But surely we should check that these assumptions hold before making them. Yet it does not seem to be possible to do this on the operationalist position. According to operationalism, we can only use a concept after it has been given an operational definition. So we have to check that space is Euclidean for distances of the order of a few kilometres *before introducing the concept of length.* Surely this is impossible!

A related difficulty is concerned with the improvement of methods of measurement. Suppose we introduced a naive definition of length in terms of rigid metre rods and employed it to measure lengths up to say half a kilometre. Then the theodolite method is discovered. At once it is employed for lengths of more than 50 metres. Now normally we would say that a *more accurate* method of measuring lengths more than 50 metres had been discovered. On the operationalist view, however, this manner of speaking is inadmissible. We have *defined* length by the rigid metre rod procedure, and the most we can say of another method of measurement is that it gives results in approximate agreement with the defining procedure for length. It makes no sense to say that the results given by the alternative method are nearer to the true value of the length than those given by the defining method. That would be like first *defining* a metre as the distance between two marks on this rod, and then saying that more accurate measurement has revealed that the distance was not exactly a metre.

A further problem for the operationalist is posed by the fact that nearly all methods of measurement have to be subjected to corrections. Consider again our simple definition of length in terms of rigid metre rods. For this to be viable in practice we will often have to make several corrections. Thus we will have to make sure that the temperature of the rod is the standard temperature for which its length was defined, or else introduce a temperature correction. If the rod is used to measure a vertical distance, we may have to correct for gravitational distortion. Then again, if the rod is made of iron, we may have to correct for electrical or magnetic distortions. Let us now consider the question of temperature corrections in rather more detail. Suppose we had defined length in terms of a measuring procedure using a metal rod but without taking temperature corrections into account. One day bright sunshine falls through the windows of the laboratory heating both the measuring rod and the wooden block being measured. It is observed that relative to the rod the wooden object has changed its length from the day before (in fact contracted). However, an intelligent experimenter then suggests that in fact the measuring rod has expanded more than the wooden block. He cools down the rod to normal room temperature and produces a more correct value of the new length of the wooden block. Indeed he now shows that it has expanded rather than contracted. But how is this admissible on the operationalist point of view? Length has been defined by the initial set of procedures and according to this definition the block must have contracted rather than expanded.

The only line that the operationalist can take on all this is to say that we have decided to adopt a new definition of length. Our naive rigid metal-bar definition of length is replaced for distances over 50 metres by a theodolite definition, whereas in certain other circumstances a temperature correction is introduced. But the operationalist now has to give an account of how new definitions are evolved and why we choose to adopt one definition rather than another.

The problems here for the operationalist become even worse if we remember that the concepts involved in the corrections must *themselves* be operationally defined. Does this not lead to a vicious circle? Popper argues that it does:

> Against this view [operationalism], it can be shown that *measurements presuppose theories.* There is no measurement without a theory and no operation which can be satisfactorily described in non-theoretical terms. The attempts to do so are always circular; for example, the description of the measurement of length needs a (rudimentary) theory of heat and temperature-measurement; but these, in turn, involve measurements of length.
>
> (1963:62)

I can see no way out of these difficulties for the operationalist, and I will therefore now turn to expounding an alternative theory of conceptual innovation in the natural sciences which does resolve the problems.

The basic idea of this theory of conceptual innovation is that the new concepts are introduced not by operational definitions, but as undefined terms in a theory, which are partially characterised by the assumptions of the theory. The theory is then brought into relation with observation by attempting to derive from it observational facts or laws. In these derivations qualitative assumptions regarding the new concept or concepts may be made. If satisfactory explanations of the observed facts or laws are obtained, the theory is regarded as confirmed, and it may then be used to devise methods for measuring the values of the new concepts in particular situations.

I will now illustrate this account of conceptual innovation with the example of Newtonian mass. The example is an appropriate one, for the concept of mass (as opposed to weight) effectively did not exist before Newton put forward his new theory of mechanics.[2] Many relevant results had indeed been established by observation and experiment, notably Galileo's law of falling bodies and Kepler's laws of planetary motion, but these results could be stated without using the concept of mass. This then is a convenient case for examining how a new concept is brought into relation with observational and experimental findings. In order to simplify the discussion I will confine myself to considering how Kepler's third law was derived from Newton's theory.

Kepler's laws are concerned with the motion of a planet P round the Sun S (Figure 7.1).

According to Kepler's first law the path of P is an ellipse with the Sun S at one focus. Suppose the length of the major semi-axis of the ellipse is a, and the period of P's orbit is T, then Kepler's third law states that a^3/T^2 is constant for all planets.

Kepler's laws do not involve the concepts of force or mass, but these concepts were introduced by Newton into his theory, which can be summarised by the familiar vector equations:

$$\mathbf{P} = m\mathbf{f}$$

and

$$\mathbf{F} = \frac{\gamma m_1 m_2}{r^2}\hat{\mathbf{r}} \quad \text{(the law of gravity)}$$

Let us suppose that the planet P has mass m_P, and that the Sun S has mass m_S. Let us neglect the gravitational interactions between the planets themselves, thus reducing the problem to a two-body problem. If we then apply the equations of Newton's theory, we can deduce[3]

$$a^3/T^2 = \gamma\,(m_S + m_P)/4p^2$$

We now assume that the mass of the Sun is very much greater than that of the planet ($m_S \gg m_P$) and so obtain

$$a^3/T^2 \approx \gamma\, m_S/4p^2 \text{ (i.e. constant)}$$

This is an approximate version of Kepler's third law. The assumption $m_S \gg m_P$, though automatically made and easily overlooked for this reason, contains the solution to the problem we have been discussing. Do we need an operationalist definition of mass at this point? Not at all. We test out theory involving masses by

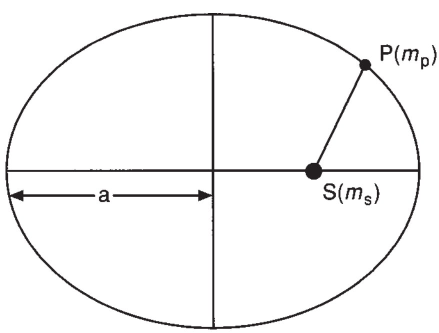


Figure 7.1 A planet P, mass m_P, moving round the Sun S, mass m_S

making the qualitative physical assumption that one mass is very much greater than another. Moreover, this qualitative assumption is justified by a crude (or intuitive) notion of mass. If we think of mass as 'quantity of matter', then observing that the Sun is very much larger than the planets and making the reasonable postulate that the density of its matter is at least similar to that of the matter in the planets, we obtain that m_S » m_P. So we do not at first need a precisely defined notion of mass. A rather crude and intuitive notion of mass can lead to a qualitative assumption and so to a precise test of a theory involving an exact idea of mass.

This example shows how a new theory involving new concepts can explain observations and observational laws not involving these concepts without the need for an operational definition of the new concepts. Newton's new theory in fact explained a great deal. Besides Galileo's and Kepler's laws, it was able to explain the laws of impact, the tides and the inequalities of the Moon's motions, and it also permitted the derivation of theories of the figure of the Earth and of comets which could be checked against observation. It also could explain certain deviations from Kepler's laws, for example perturbations in the orbits of planets caused by gravitational interactions between the planets. All these explanations and derivations could be analysed in much the same way as we have done for Kepler's third law, and the general conclusion is that Newton's new theory could be tested out against observational data and receive considerable confirmation without the need for any operational definition of the new concepts it contained. Let us therefore pass to the second stage of our account of conceptual innovation in the exact sciences. Suppose a new theory with new concepts has been tested and confirmed, it can then be used to obtain measurements of the new concepts in particular situations, again without the need for any operational definitions. I will now illustrate this procedure in the case of Newtonian mass.

Let us therefore consider how the mass of a planet such as Mars could be measured. Of course this example is deliberately chosen to illustrate the uselessness of introducing operational definitions in terms of experimental procedures with balances. Such an approach will get nowhere, but, by making subtle and appropriate calculations from the theory, the problem can be solved. Let us suppose therefore that we have a planet P which is distant a_P from the Sun and has orbital period T_P. Then, assuming m_S » m_P, we have as before

$$\mathrm{a}_P{}^3/T_P{}^2 \approx \gamma m_S/4\mathrm{p}^2$$

But now suppose that the planet has a moon M of mass m_M which is distant a_M from the planet and has orbital period T_M. If we assume that m_P » m_M, we get as before

$$\mathrm{a}_M{}^3/T_M{}^2 \approx \gamma m_P/4\mathrm{p}^2$$

Therefore dividing, we obtain

$$m_P/m_S \approx (\mathrm{a}_M/\mathrm{a}_P)^3\,(T_P/T_M)^2$$

All the quantities on the right hand side of this equation can be determined by astronomical measurement and so we obtain a value for the ratio m_P/m_S. For example, Mars has a moon Deimos, whose period is 30.3 hours. We hence obtain $m_{Mars}/m_{Sun} = 3.4 \times 10^{-7}$, which gives a measurement of the mass of Mars relative to that of the Sun.

We are now in a position to state our non-operationalist theory of conceptual innovation in the natural sciences. Let us suppose that a new theory is proposed involving new concepts. We first test the new theory by deducing from it consequences which do not involve the new concepts and comparing these consequences with experience. To make these deductions we do not need any operational definitions of the new concepts, but, regarding these new concepts, we will in general make some qualitative assumptions of approximate equality or great inequality in particular physical situations, such as the assumption that $m_S \gg m_P$. The new theory together with these qualitative assumptions will lead to the conclusion that some consequences hold approximately. These consequences can then be matched against the results of experiments or observations past or future. If the new theory is corroborated by these comparisons, it is accepted and methods for measuring values of the new concepts in particular situations are devised on the basis of it, as we illustrated by the example of measuring the mass of Mars relative to that of the Sun. Once again, no operational definitions are needed at this stage of devising methods of measurement.

I will now try to show that this theory of conceptual innovation avoids the difficulties in operationalism which we noted earlier. The first problem arose because new operational definitions are needed as a concept is extended into new fields. The laws on which these new operational definitions are based must apparently be checked before introducing the concept itself. This was illustrated by the example of extending the rigid metre rod definition of length by using a theodolite. The theodolite is based on Euclidean geometry whose truth must apparently, on the operationalist position, be checked before introducing the concept of length.

Our theory resolves this problem because it takes concepts as acquiring meaning *not* through operational definitions, but through their position in a nexus of theories. An account of the logical relations of these theories and of the way we handle them in practice gives us the significance of the concept. Thus a concept can indeed be extended, *not* by acquiring new operational definitions, but rather by becoming involved in a series of new and more general theories. If we accepted the operationalist view, we could not suddenly postulate a new theory with new concepts. The new concepts would only have meaning after they had been operationally defined. An operationalist must therefore check the laws on which his definitions are to be based *before* introducing the concept. In general, this programme cannot be carried through as can be seen from the absurdity of trying to check whether Euclidean geometry holds before introducing the concept of length. Moreover, our non-operational theory of conceptual innovation shows that it is unnecessary. We are quite free to introduce a new undefined concept in a new theory. The only problem then is how to test this theory, and this problem, as we have seen, can be solved by making qualitative assumptions regarding the new concept in particular situations.

A further difficulty in operationalism was the question of how the operationalist could give an account of the correction and improvement of methods of measurement. We often, for example, speak of 'discovering a more accurate method of measuring a concept', but if the previous method was the *definition* of the concept how is any more accurate method of measuring it possible? Again we often introduce corrections for temperature, gravitational forces, etc. But how can we correct a definition? And it we try to do so, does this not lead to a vicious circle in which, for example, length is defined in terms of temperature, and vice versa?

All these difficulties disappear as soon as we recognise the primacy of theories. Methods of measurement are only introduced on the basis of theories; and there is no reason why, starting from a particular set of theories, we should not be able to devise two methods of measurement – one more accurate than the other. Again, our methods of measurement involve not only the general theories but also certain qualitative assumptions, e.g. that temperature variations in the laboratory are negligible. We can always replace such an assumption by a more sophisticated one, thus correcting our previous method of measurement.

This concludes my criticisms of operationalism and exposition of a non-operationalist theory of conceptual innovation in the natural sciences. In the next section I will attempt to apply this theory to probability in order to produce my version of the propensity interpretation. Before doing so, it will be useful to introduce one further concept which will be helpful in our development of the propensity theory of probability. This is the concept of the *depth* of a scientific theory.

Popper (1957c)[4] introduces this concept of depth where he writes:

> Every time we proceed to explain some conjectural law or theory by a new conjectural theory of a higher degree of universality, we are discovering more about the world, trying to penetrate deeper into its secrets.
>
> (1957c:28)

Popper does not attempt to give a full account of the sense in which one theory can give a *deeper* description of reality than another. He does however suggest a sufficient condition for a higher level theory to have greater depth than a lower level theory which it explains. This is illustrated by the historical example of Newtonian mechanics.

Popper observes that Newton's theory does not just explain Galileo's and Kepler's laws, but he also corrects them. For example, Kepler's first law states that all planets move in ellipses with the Sun at one focus. The approximate truth of this does indeed follow from Newton's theory, but Newton's theory also predicts that there will perturbations of the elliptical orbits due to gravitational attractions between the planets. Generalising from this example, Popper suggests that a higher-level theory should be regarded as *deeper* than the theories it explains if it *'corrects them while explaining them'* (1957c:33). In particular Newton's theory, because it corrects Kepler's laws while explaining them, is *deeper* than Kepler's laws.

I certainly favour adopting this suggestion of Popper's, but it should be emphasised that Popper's intention was just to give one sufficient condition for depth, and he himself says that there may be others. I will therefore propose a second sufficient condition for depth, which is a slight modification of Popper's but which is more readily applicable to probability theory. Let us suppose that before Newton we did not have Kepler's law but only Schnorkelheim's law S, which stated that planets move round the Sun in closed curves which are vaguely though not exactly circular. Now Newton's theory is invented and from it we infer as usual S′: planets move round the Sun in ellipses which are disturbed by small perturbations. Now S′ does not contradict S in the way that it contradicts Kepler's first law. In fact S′ entails S. Thus S′ does not correct S, but it does *render* S *more precise.* This 'rendering of a law more precise' also seems to me a sufficient condition for greater depth. So I would extend Popper's condition to the following: 'a higher-level theory has greater depth than a lower-level one if it *corrects* or *renders more precise* the older theory while explaining it.'

Let me now briefly sketch how the ideas of this section are going to be applied to probability theory. In mechanics a set of empirical laws, notably Galileo's laws and Kepler's laws, were explained by a new deeper theory (Newton's theory) which involved the new and undefined concept of 'mass'. In probability theory too we have empirical laws: the Law of Stability of Statistical Frequencies and the Law of Excluded Gambling Systems. Our aim should be to exhibit probability theory as a theory which explains these laws in terms of the new concept of probability. Further, probability theory should not only explain these laws, but also correct them or render them more precise – thus proving itself to be a deeper theory. The new concept 'mass' did not acquire empirical significance through an operational definition, but through the assumption that one mass (the mass of a planet) was negligible in comparison with another (the mass of the Sun). If the analogy is going to hold here too, probability will not acquire empirical significance by means of a definition in terms of relative frequency (as Von Mises claimed), but *through the decision to neglect one probability in comparison with another.* In the rest of the chapter a development of probability theory along the lines just indicated will be attempted.

A falsifying rule for probability statements

Our problem is how exactly the theoretical term (probability) is linked to the observable term (frequency). It turns out the solution to this problem lies in the consideration of another problem which Popper posed for the philosophy of probability. This is the question of how falsifiability applies to probability. Having advocated falsifiability in *The Logic of Scientific Discovery,* and having also a considerable interest in probability, it was very natural for Popper to consider how falsifiability applies to probability and this is just what he does do in Chapter VII of his famous work. It turns out that there is a difficulty connected with the falsifiability of probability statements which Popper himself states very clearly as follows:

> The relations between probability and experience are also still in need of clarification. In investigating this problem we shall discover what will at first seem an almost insuperable objection to my methodological views. For although probability statements play such a vitally important rôle in empirical science, they turn out to be in principle *impervious to strict falsification.* Yet this very stumbling block will become a touchstone upon which to test my theory, in order to find out what it is worth.
>
> (1934:146)

To see why probability statements cannot be falsified, let us take the simplest example. Suppose we are tossing a bent coin, and we postulate that the tosses are independent and that the probability of heads is p. Let Prob(m/n) be the probability of getting m heads in n tosses. Then we have

$$\mathrm{Prob}(m/n) = {}^{n}C_{m}\, p^{m}\, (1 - p)^{n - m}$$

So, however long we toss the coin (that is however big n is) and whatever number of heads we observe (that is whatever the value of m), our result will always have a finite, non-zero probability. It will not be strictly ruled out by our assumptions. In other words, these assumptions are ‘in principle *impervious to strict falsification.*’

Popper’s answer to this difficulty consists in an appeal to the notion of methodological falsifiability. Although, strictly speaking, probability statements are not falsifiable, they can nonetheless be used as falsifiable statements, and in fact they are so used by scientists. He puts the matter thus: ‘... a physicist is usually quite well able to decide whether he may for the time being accept some particular probability hypothesis as “empirically confirmed”, or whether he ought to reject it as “practically falsified” ...’ (Popper 1934:191).

Popper’s approach has been strongly vindicated by standard statistical practice. Working statisticians are constantly applying one or other of a battery of statistical tests. Now, whenever they do so, they are implicitly using probability hypotheses, which from a strictly logical point of view are unfalsifiable, as falsifiable statements. The procedure in any statistical test is to specify what is called a ‘rejection region’, and then regard the hypothesis under test (H say) as refuted if the observed value of the test statistic lies in this rejection region. Now there is always a finite probability (called the ‘significance level’ and usually set at around 5 per cent) of the observed value of the test statistic lying in the rejection region when H is true. Thus H is regarded as refuted, when, according to strict logic, it has not been refuted. This is as much as to say that H is used as a falsifiable statement, even though it is not, strictly speaking, falsifiable, or, to put the same point in different words, that methodological falsifiability is being adopted.

The first important statistical tests were introduced in the period 1900–35. Karl Pearson (1900) proposed the chi-square test, and W. S. Gosset (1908), who modestly wrote under the name 'Student', introduced the *t*-test. It was at this point that Fisher began his work. He gave a better mathematical foundation to the tests of Karl Pearson and Student, and introduced his own *F*-test and the analysis of variance. Two of Fisher's books were important for the introduction of these new ideas and techniques to statisticians. The first was *Statistical Methods for Research Workers* (1925), and the second was *The Design of Experiments* (1935). The chi-square test, *t*-test and *F*-test are still widely used today, though other tests have of course been devised subsequently. Now the interesting thing is that statistical tests were introduced and came to be very widely adopted quite independently of Popper's advocacy of methodological falsifiability. Statistical tests are, however, based implicitly on methodological falsifiability, and their introduction and widespread adoption by statisticians provides striking corroboration of the value of Popper's approach.

Let us now try to formulate methodological falsifiability as applied to probability a little more precisely. The idea of methodological falsifiability is that, although probability statements are not strictly speaking falsifiable, they should be used in practice as falsifiable statements. If we adopt this position, we ought to try to formulate what could be called a *falsifying rule for probability statements* (FRPS) which shows how probability statements should be used as falsifiable statements. Such a rule should obviously agree with, and implicitly underlie, the practice of statistical testing. So modelling our rule on some of the standard statistical tests we obtain an FRPS which can be stated roughly as follows.

Let H be a statistical hypothesis, and suppose we are trying to test H against some evidence which consists of a sample of n data points $\{e_1, e_2, ..., e_n\}$. Let X be a test statistic, that is to say a function $X(e_1, e_2, ..., e_n)$ of the observed data whose value can be calculated from the data. Suppose that we can repeatedly and independently draw such samples of size n, then X is a random variable, which takes on different values for the different samples. Suppose that from H it can be deduced that X has a bell-shaped distribution D of roughly the form shown in Figure 7.2.

We shall call distributions of this shape *falsifiable distributions.* Two points a and b are chosen so that D is divided into a 'head', i.e. $a \leq X \leq b$, and tails, i.e. $X < a$, or $X > b$. The tails are such that the probability of obtaining a results in the tails, given H, has a low value known as the significance level. The significance level is normally chosen between 1 per cent and 10 per cent, 5 per cent being the most common value. Our falsifying rule for probability statements now states that if the value obtained for X is in the tails of the distribution, this should be regarded as falsifying H; whereas, if the value of X is in the head of the distribution, this should be regarded as corroborating H. Informally the FRPS can be characterised as that of cutting off the tails of falsifiable distributions.

Broadly speaking, this falsifying rule agrees with the practical procedures adopted when such standard statistical tests as the chi-square test, the *t*-test or the *F*-test are applied. There are indeed some small divergences connected with the

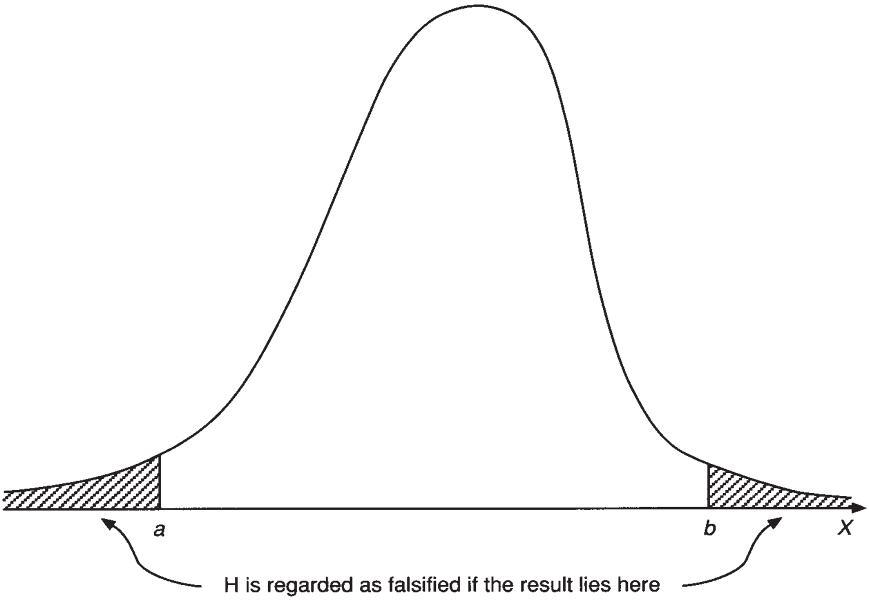


Figure 7.2 The falsifying rule for probability statements (FRPS)

use of one-tailed as opposed to two-tailed tests, and a problem posed by the Neyman paradox. These are rather technical matters, however, and I will not deal with them here, though the interested reader will find a list of fuller and more mathematical treatments of the question in Note 5 to this chapter.[5] In this book I will simply draw attention to the undoubtedly broad agreement between the proposed falsifying rule and the practice of statistical testing, and make just one further observation in favour of the rule.

This observation is that our falsifying rule for probability statements is a very natural generalisation of the way in which errors are treated in deterministic science. I will try to show this by relating the problem to that of 'intervals of imprecision', a course followed by Popper (1934: §68). Suppose we are testing a deterministic hypothesis H. We might in some simple cases deduce that, given H, a particular measurable quantity x should have a value x_0. We would then measure x to see whether it did indeed equal x_0 or not. More usually, we might deduce, given H, that two measurable quantities y and z were linearly related $y \propto z$. We would then measure a number of pairs of values of y and z, (y_1, z_1), ..., (y_n, z_n) say, and see whether these pairs did indeed fall on a straight line. Now would we regard H as falsified if x differed by any quantity, however small, from x_0, or if the curve joining the (y_i, z_i) departed in any degree whatsoever from linearity? Of course not. Indeed we would expect x to differ from x_0 to some extent, and would be surprised if the two quantities were experimentally indistinguishable or if the (y_i, z_i) lay *exactly* on a line. The difference from x_0 or the departure from linearity would be attributed to 'experimental error'. But now could we say that the experiment agreed with H however much x differed from x_0, or however randomly the points (y_i, z_i) were scattered on the plane? Could we argue that the experimental errors had just been large in such cases? Once again we would of course reject such an absurd view. In fact we would regard the hypothesis as confirmed if we obtained results sufficiently near x_0 and falsified if the result were too far away. In other words, we would surround x_0 by an approximately defined 'interval of imprecision' $[x_0 - \phi, x_0 + \phi]$ and regard H as confirmed if the result were in the interval and falsified otherwise. Similarly, in the (y, z) case we would take a band in the plane which we would regard as a sufficient approximation to a straight line. Now this procedure is surely very similar to the adoption of a falsifying rule. In both cases we could in theory allow any divergence, however large. Yet in practice we draw the line in an admittedly somewhat arbitrary fashion and only permit divergences up to a certain point. In both cases this decision makes the underlying hypothesis falsifiable.

The analogy is heightened if we consider that experimental errors can be treated statistically. Let us show this in the simple case in which we predict $x = x_0$. The (y, z) example is similar but involves considerations of regression. Various measurements of the quantity concerned can be considered as independent trials. These results of these trials are given as values of $X = x - x_0$, where X is a random variable. The value of X on a particular trial shows the degree of error in the measurement, i.e. the degree to which it deviates from the predicted value of x_0. If we now assume that this error is the sum of a large number of mutually independent elementary errors, we obtain that X is distributed approximately normally about the value 0. Of course the assumption behind this deduction is not always very plausible, and indeed other distributions of the error random variable often agree better with observation.

The actual form of the distribution is not important for our purposes however. The question is this. Suppose we assign some distribution D to the error random variable X. How does this more sophisticated statistical treatment tie in with the usual procedure of assigning an interval of experimental imprecision $[x_0 - \phi, x_{0\,+}\, \phi]$ to x_0? The answer is that D must be a falsifiable distribution whose head (or acceptance region) A is the interval of imprecision. But now we see that the selection of an interval of imprecision and the application of the falsifying rule for probability statements become in this case exactly equivalent. Thus the adoption of the FRPS can be seen as a natural generalisation of the way in which errors are treated when testing deterministic theories. Indeed we could formulate the difference between statistical and deterministic science in the following fashion. In deterministic science no statistical considerations appear in the laws and theories themselves and only come in when these laws and theories are tested. In statistical science probability enters the laws as well.

That concludes my account of the proposed falsifying rule for probability statements (FRPS). I must next explain the rôle which this rule plays in the version of the propensity theory of probability to be developed here. In the frequency theory the link between probability and frequency was established by giving an operational definition of probability in terms of frequency. In the present version of the propensity theory the link is established instead by adopting the falsifying rule. With the help of the FRPS, we can derive from probability hypotheses results about frequencies, and these can be checked by observation. In particular, from the axioms of probability (in a suitable formulation) we can derive the two empirical laws to which Von Mises drew attention – that is, the Law of Stability of Statistical Frequencies and the Law of Randomness. This is analogous to the way in which Kepler's and Galileo's laws were derived from Newton's theory. In the Newtonian case the derivation was accomplished by neglecting one mass (the mass of a planet) in comparison with another mass (the mass of the Sun). In the probability case the derivation will be accomplished with the help of the falsifying rule, but this can be considered as equivalent to neglecting one probability (the probability of getting a result in the tails of the distribution) in comparison with another (the probability of getting a result in the head of the distribution). The analogy here is really quite close, but there are differences between the two cases as we shall see when we look in more detail at the derivation of the empirical laws of probability in the next section.

Derivation of the empirical laws of probability

Our aim is to show how, from the axioms of probability in a suitable formulation, we can, with the help of the falsifying rule, derive the two empirical laws of probability. Although the derivation in its full generality is not difficult, it does require some mathematical background. I will therefore give the derivation in this section in a simple special case. In the next section which is one of the starred (or mathematical) sections, I will discuss the nature of the axioms in this version of the propensity theory. It will be obvious in the light of this discussion how the derivation of the present section can be generalised.

Let us therefore take the simple case of tossing a possibly biased coin. Let us suppose that the tosses are independent, and the probability of heads is p. I will begin by deriving the Law of the Stability of Statistical Frequencies for this case using our falsifying rule. Now the mathematics of coin tossing was well worked out in the eighteenth century, and is summarised in note 4 to Chapter 1 (pp. 206–7). The probability of getting r heads in n tosses is given by the *binomial* distribution

$$\text{Prob}(r \text{ heads in } n \text{ tosses}) = {}^{n}C_{r}\, p^{r}(1 - p)^{n-r}$$

This is a discrete distribution, but for large n it tends to a continuous distribution known as the *normal* or *Gaussian* distribution, whose formula is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Figure 1.1 gives an illustration of how the binomial distribution tends to the normal distribution. It will be seen that for n as low as 30 the approximation is a very good one. All we have to do now is apply our falsifying rule to the normal distribution by 'cutting off its tails', and this will enable us to draw conclusions about frequencies.

The binomial distribution as given above has mean p and standard deviation

$$\sqrt{\frac{p(1-p)}{n}}$$

It is convenient to consider the standardised variable

$$X = \frac{r/n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

The distribution of X tends to the normal distribution with zero mean ($\mu = 0$) and unit standard deviation (s = 1) as n ? 8. If we apply our falsifying rule to the normal distribution with mean 0 and unit standard deviation, using a significance level of 5 per cent, the tails of the distribution turn out to correspond to values of the random variable greater that +1.96, and less than -1.96. So, using the normal approximation to the binomial distribution, we infer that with a probability of 95 per cent that

$$-1.96 \leq X \leq +1.96$$

$$p - 1.96\sqrt{\frac{p(1-p)}{n}} \leq \frac{r}{n} \leq p + 1.96\sqrt{\frac{p(1-p)}{n}} \tag{7.1}$$

Adopting our falsifying rule is tantamount to regarding Equation 7.1 as practically certain, and this completes our derivation, for Equation 7.1 is just a form of the Law of Stability of Statistical Frequencies. It says that as $n \to \infty$ the observed frequency r/n will tend to a fixed value p. Moreover it improves on the rough empirical statement of the law by telling us that the rate of convergence is of the order of $1/\sqrt{n}$

Let us now consider the case of an unbiased coin for which $p = 1/2$. In this case Equation 7.1 becomes

$$\frac{1}{2} - \frac{0.98}{\sqrt{n}} \leq \frac{r}{n} \leq \frac{1}{2} + \frac{0.98}{\sqrt{n}} \tag{7.2}$$

So, if we apply our falsifying rule at the 5 per cent level, we, in effect, predict that it is practically certain that r/n will lie in the interval $\pm 0.98/\sqrt{n}$ and regard our underlying hypothesis as refuted if the observed value of r/n lies outside this interval.

To see how this applies in practice, I give in Table 7.1 the results of some coin-tossing experiments. The first was performed by me, the second by Buffon, and the third and fourth by Karl Pearson. In each case I give the allowable deviation around 0.5 as calculated by the falsifying rule at the 5 per cent level of significance, and also the observed relative frequency of heads and its actual deviation from 0.5. As can be seen, the results of all four experiments confirmed the hypothesis of an unbiased coin and independent tosses. These results show in a vivid way that, even if probability is not defined in terms of frequency, the adoption of a falsifying rule for probability statements can establish a link between probability and observed frequency.

The analogies between this situation and that of Newtonian mechanics in relation to Kepler's laws are clear. As we have shown, Kepler's third law was obtained from Newton's theory by neglecting one mass (the mass of a planet) in comparison with another mass (the mass of the Sun). Similarly the Law of Stability of Statistical Frequencies is obtained from probability theory by neglecting one probability (the probability of a result in the tails of the distribution) in comparison with another probability (the probability of a result in the head, or acceptance region of the distribution). There is moreover a further point of similarity. We have given Popper's view that Newton's theory showed itself to be deeper than Kepler's laws because it corrected them while explaining them. In particular, Kepler's first law states that all planets move in ellipses, and this is corrected by Newton's theory to the law that planets move in ellipses which are disturbed by small perturbations caused by gravitational interactions between the planets. Now the empirical Law of Stability of Statistical Frequencies is not corrected by probability theory because it is rather a vague law. It states that the observed frequency will tend towards a stable value for large n, but does not give the rate of convergence or even rough limits on the possible divergences for different values of n. Probability theory does not therefore correct the law, but it does *render it more precise,* and this, as I argued earlier, is another perfectly good reason for regarding one theory as deeper than another which it explains. As we have seen, the calculations of probability theory tell us

Table 7.1 Some coin-tossing experiments

Author	*Number of tosses*	*Allowable deviation*	*Relative frequency of heads*	*Difference between relative frequency and 0.5*
Gillies	2,000	±0.022	0.487	–0.013
Buffon	4,000	±0.015	0.507	+0.007
Karl Pearson	12,000	±0.009	0.502	+0.002
Karl Pearson	24,000	±0.006	0.501	+0.001

that the rate of convergence is of the order of $1/\sqrt{n}$, and they also give approximate limits to the allowable divergences for different values of n. This certainly makes the law more precise, and so probability theory shows itself to be deeper than the empirical Law of Stability of Statistical Frequencies which it explains. Our version of the propensity theory therefore justifies a claim made by Miller when he writes: 'One of the virtues of the propensity interpretation of probability is that it offers a somewhat deeper explanation of statistical stability.' (1996:138).

The account just given overcomes some of the difficulties which we noted in Von Mises' treatment of the Law of Stability of Statistical Frequencies. As we saw earlier (pp. 93–5), Von Mises wanted to give a more precise statement of this law, but, being an operationalist, this had to be obtained simply from empirical investigations before introducing the concept of probability. Probability would then be defined using an axiom obtained by abstraction from the empirical law. Yet this order of events, as we argued previously (pp. 94–5), was neither accurate historically nor feasible in practice. The Law of Stability of Statistical Frequencies might well start as a rough empirical law, but it could not have been made precise without introducing the theoretical concept of probability and relating this concept to observed frequencies in something like the way that we have explained. It is a pure fiction to claim that the law could have been made more precise by enormously long and complicated investigations of coin tossing which were not guided by any theoretical ideas. Our view, which emphasises the continual interaction between theory and observation, is simpler and more practical than the operationalist strategy of trying to do all the observing first, and only then introducing the theoretical concepts.

The above account also avoids all the difficulties which Von Mises faced concerned with the approximation of the large finite collectives observed in practice by the infinite collectives postulated in the theory. Using our falsifying rule we can handle empirical collectives of length n, where n can be 2,000, 4,040, 12,000, 24,000, or indeed any other definite number. There is no need to say that we approximate, for example, 24,000 tosses by an infinite sequence of tosses. A supporter of Von Mises might reply to this that the above derivation does also approximate the large finite to its limit at infinity, though at a different point. After all, in the derivation the binomial distribution is replaced by the normal curve to which it tends in the limit. This is true of course, but I would maintain that this use of limits is quite unproblematic. It is not a question of relating some empirical reality to a hypothetical mathematical limit, but rather it is the use of a limit as a mathematical approximation for computational purposes. It is a purely mathematical matter whether, and to what extent, the binomial distribution for large n approximates to the normal distribution. We can estimate the degree of the approximation mathematically. Indeed in a particular case we could dispense with it altogether and use a computer to work out the exact values of the binomial distribution.

Let us now turn to the Law of Randomness or of Excluded Gambling Systems. In the special case which we are considering, this law states that in any subsequence of tosses selected from the original sequence by a gambling system, the observed frequency will still approximate to p. The derivation of this law is very easy. Suppose we select a subsequence from the original sequence by means of a gambling system. This subsequence will still be a sequence of independent tosses for which Prob(heads) = p. So we can apply the mathematical analysis given above to the subsequence. If it is of length n, we can conclude as before that, at the 5 per cent significance level, Equation 7.2 above will hold, which tells us that the observed frequency for the subsequence will approximate to the probability p, that is to the same value as for the sequence of tosses as a whole.

The above derivation shows that in the propensity theory the notion of randomness is really reduced to that of independence, and indeed we can define random sequences in terms of independence as follows. Let us confine ourselves to sequences of 0s and 1s. We shall say that such a sequence is random if it is generated by repeating a set of conditions **S** which are such that (a) the repetitions are independent, and (b) the outcomes are 0 or 1 with Prob(0) = p, for some fixed value p such that $0 \leq p \leq 1$. This definition includes the cases 00 ... 0 ... and 11 ... 1 ... as degenerate random sequences. This consequence also held in Von Mises' approach and is harmless. It is worth noting that the definition introduces a great mathematical economy relative to the frequency approach. In the frequency theory, random sequences were defined by the invariance of limiting frequencies with respect to a set of gambling systems. However, when it came to considering the combination of collectives, Von Mises had to define and use a notion of *independent* collectives. Thus he introduced two ideas, 'randomness' and 'independence', which were quite differently defined, although it is clear that these two notions are really one and the same. This fact is shown in the above definition which in effect reduces the notion of randomness to that of independence, and thereby simplifies the mathematical development.

But what rôle do gambling systems now play in the theory? The answer is a simple one. They can be used to obtain tests of independence. This can be illustrated by the results of a simple experiment which I carried out some years ago. It consisted of tossing an old penny 2,000 times and noting the sequence of heads and tails obtained. The hypothesis was that the tosses were independent with Prob(heads) = $^1/_2$. As noted in Table 7.1, the observed frequency of heads in the 2,000 tosses was 0.487 whose difference from 0.5, i.e. -0.013, was within the deviation ±0.022 allowed by the falsifying rule at the 5 per cent significance level. This observation could be considered as a test particularly directed at the assumption that the coin was unbiased. It is surely desirable, however, that this test should be supplemented by others more specifically directed at the assumption that the tosses were independent. Now tests of independence could be obtained by selecting some subsequence of the original sequence of 2,000 tosses using a gambling system, and checking that the relative frequency of heads in this subsequence still satisfies the relation given in Equation 7.2 above. Altogether ten gambling systems were employed. First of all, every second toss was selected [g(2)]. This system could be started at the first toss [g(21)] or at the second [g(22)]. Next, every fourth toss was selected. This gave in a similar way four gambling systems g(41), g(42), g(43) and g(44). Then the sequences of results were noted which followed a single head [g(AH), standing for g(After Head)], a single tail [g(AT)], two heads [g(AHH)] and finally two tails [g(ATT)]. For each gambling system we record in turn (see Table 7.2) the number of members of the corresponding subsequence, the deviation of the relative frequency from 0.5 which by Equation 7.2 is allowable, the observed relative frequency of heads and the difference between the observed relative frequency of heads and 0.5. If the observed difference is within the allowable deviation the hypothesis is confirmed. If not, it is falsified.

Examining Table 7.2, we see that 10 of the 11 tests confirm the hypothesis, but one of them [g(43), marked with an asterisk in Table 7.2] gives a falsification. The allowable deviation is ±0.044, whereas the actual deviation is -0.048, just outside the interval. In these circumstances what should we take to be the overall result? At this point it is worth drawing attention to a peculiar feature of the FRPS. Suppose we adopt a significance level of 5 per cent. This means that if we subject a true statistical hypothesis to a battery of tests, we should expect to have an erroneous falsification in about one case in twenty. It follows that our falsifying rule must be used with a certain 'judiciousness'. If a particular test results in a falsification, we cannot automatically assume that the hypothesis should be regarded as refuted. In some cases this would be a reasonable conclusion to draw, but in other cases the overall picture would tell a different story. In the present case, the falsifying test is one of a group of eleven tests and the others all give confirmations. This is in a situation where, if the hypothesis is indeed true, we would expect one in twenty of the tests made to give an erroneous falsification. Moreover, the observed result is only just outside the allowed interval. Indeed at a significance level of about 2.8 per cent or less, the test would have given a confirmation. Putting all these considerations together, the natural conclusion is that the tests give an overall

Table 7.2 Gambling systems in a coin-tossing experiment

Gambling system	*Number of observations*	*Allowable deviation*	*Relative frequency of heads*	*Difference between relative frequency and 0.5*
None	2,000	±0.022	0.487	–0.013
g (21)	1,000	±0.031	0.470	–0.030
g (22)	1,000	±0.031	0.504	+0.004
g (41)	500	±0.044	0.488	–0.012
g (42)	500	±0.044	0.510	+0.010
g (43)	500	±0.044	0.452	–0.048*
g (44)	500	±0.044	0.498	–0.002
g (AH)	974	±0.031	0.505	+0.005
g (AT)	1,025	±0.031	0.470	–0.030
g (AHH)	487	±0.044	0.503	+0.003
g (ATT)	542	±0.042	0.482	–0.018

*The only falsification.

confirmation of the underlying hypothesis, i.e. that the coin was unbiased and the tosses independent.

It might, however, be objected that the need for using our falsifying rule 'in a judicious fashion' detracts considerably from its appeal. I will give another example of such a judicious use of the rule in a moment, and I will then discuss in general terms the problem posed by the need for this judiciousness. For the moment, however, I would like to point out that the situation here has some analogies with what occurs in non-statistical branches of science. The results of a single scientific test are rarely if ever conclusive. In physics a single test can reveal a 'stray effect' which never appears on subsequent repetitions of the test. A famous example of this was the positive result on the Michelson–Morley experiment obtained by Miller. As further repetitions of the experiment continued to give the original negative result, it was concluded the Miller's result must have been owing to some unknown cause of error, although the matter was never fully cleared up.

In general, if we suspect that we are dealing with a stray effect, we can always repeat the test a number of times. If the phenomenon never reappears, we would disregard it as being a mere oddity. In the same way, if we suspect that a particular application of the FRPS has given an erroneous falsification, we can always carry out a battery of statistical tests. If only one of these gives a falsification not far outside the allowed interval, and if the results of the others are all confirmations, then we can take the overall result to be a confirmation. I will now give another example of such a judicious use of the falsifying rule. It is concerned with the practical production of random numbers.

A further advantage of the definition of randomness in terms of independence given above is that it ties in very nicely with the way in which random numbers are produced in practice. An example of this procedure is provided by Kendall and Babington Smith's (1939b) tables of random sampling numbers. These authors used a kind of improved roulette wheel. It consisted of a disc divided into ten equal sections marked 0, 1, ..., 9, and rotated by an electric motor. From time to time the disc was illuminated so that it appeared to be stationary, and the number next to a fixed pointer was noted. (For a more detailed description of the randomising machine see Kendall and Babington Smith 1939a: 51–3.) A sequence of 100,000 digits was collected using this machine, and this sequence, together with some of its subsequences, was then subjected to four different kinds of statistical test.

One type of test can be considered as primarily a test of independence, although it was not based, at least in any direct sense, on a gambling system. Although gambling systems lead to tests of independence, not all tests of independence need be based on gambling systems. Kendall and Babington Smith describe this test as follows: 'The lengths of gaps between successive zeros were counted and a frequency distribution compiled for comparison with expectation. This test is called the *gap* test.' (1939b: viii).

Another kind of test used by Kendall and Babington Smith was a simple frequency test which consisted in comparing the observed relative frequencies in a subsequence with their expected values, i.e. $^1/_{10}$. The whole sequence was arranged first in 100 blocks of 1,000 digits, then in twenty blocks of 5,000 digits and finally in four blocks of 25,000 digits. Three of the tests were applied to each of the blocks of 1,000 digits and the four tests to the remaining blocks and to the whole sequence. Out of these 400 tests there were only six failures. Four of the blocks of 1,000 digits failed to pass one of the tests and one failed to pass two tests. Even here, however, the divergence from expectation was not very great.

These failures did not cause Kendall and Babington Smith to reject the hypothesis of randomness. Rather they implicitly used the falsifying rule in a judicious fashion and argued that in such a large number of tests it is likely from the nature of statistical testing that there will be a few failures. Nearly all the evidence supports the randomness assumption, and the anomalies can therefore be dismissed as stray effects. This reasoning seems entirely valid and in agreement with our earlier discussion. Kendall and Babington Smith also add that the blocks of 1,000 digits which failed at least one test are probably not very suitable for use in practical situations. I will return to this point in a moment. In 1955 the Rand Corporation published *A Million Random Digits.* The practical means used to obtain these were different but the underlying principles remained the same. The physical basis of the experiment in this case was the random emission of electrons in certain circuits.

It is interesting to observe that in both cases mentioned there were considerable practical difficulties in eliminating bias and dependencies. Let us consider Kendall and Babington Smith first. They took readings from the machine themselves and also got an assistant to take some readings. They found, as I have already mentioned, that their own readings satisfied nearly all tests for randomness. However, the assistant's readings showed a significantly higher frequency of even numbers than odd numbers. Kendall and Babington Smith concluded that he must have had a strong unconscious preference for even numbers, and that this caused him to misread the results. The Rand Corporation ran into troubles of a different kind. To make use of the random emission of electrons, it is necessary to amplify the signal. Now the amplifying circuits have a certain 'memory', and this is liable to introduce dependencies even if the underlying emissions are genuinely independent. These examples are highly instructive and also encouraging, for they show that our statistical tests do really enable us to detect biases and dependencies so that we can eliminate them.

It is now an appropriate moment to say a few words about a difficulty which arises when random numbers are used in practice. Suppose we have a sequence of digits which is random in the sense already defined, i.e. produced by the independent repetitions of an experiment whose results are 0, 1, 2, ..., 9 and have equal probability. If the sequence is sufficiently long then there will be a very high probability of having a subsequence of, say, 100 consecutive 0s. Indeed the whole sequence would fail a number of tests of randomness unless such a subsequence appeared. But now suppose we are using the random numbers in practice – say to obtain a random sample of size 100. The subsequence of 100 consecutive 0s would be most unsuitable. In other words, a sequence of random numbers may not be suitable for use in practice. Kendall and Babington Smith call a sequence of random numbers which is suitable for practical use a set of random *sampling* numbers. They then put the point thus: 'A set of Random Sampling Numbers ... must therefore conform to certain requirements other than that of having been chosen at random.' (Kendall and Babington Smith 1938:153). The problem is now: what are these further requirements? Our previous discussion suggests a simple answer which agrees with what Kendall and Babington Smith themselves say.

It has already been emphasised that statistical tests are always provisional and that it is always possible to reject apparent falsifications as 'stray effects' in the light of subsequent evidence. Thus it is perfectly possible to have a sequence which is in fact random but which fails a number of standard tests for randomness. The 100 consecutive 0s just mentioned would be an example of this. Consequently, it makes sense to require that a sequence should not only *be* random, but also satisfy certain standard tests for randomness. Such sequences, I claim, are the ones which are most suitable for practical use.

So far I have surveyed some of the empirical results obtained by coin tossing or the use of an improved roulette wheel. In fact there is a mass of empirical evidence of this kind obtained using coins, dice, roulette wheels and similar devices. Keynes (1921:361–6) gives a survey of some historical experiments of this character, while Iversen *et al.* (1971) give the results of a striking recent experiment involving more than four million dice throws. The outcomes of these various experiments is more or less in line with the ones we have discussed in more detail. A study of the results has revealed bias in dice and also bias in observers who sometimes have unconscious preferences for some numbers rather than others. A judicious use of the falsifying rule is usually needed. However, once we abstract from these points, the empirical results give very strong confirmation to the standard probability models. This is an excellent illustration of the thesis that probability theory is a science. The standard probability models give quite precise predictions of what frequencies should be observed, and there is, as far as I can see, no a priori or logical reason why observations should agree with these predictions. For example, even if convergence to a fixed value is observed, why should this convergence occur at the rate of $1/\sqrt{n}$? Yet convergence does indeed occur at this rate, and this confirms the basic principles of probability theory.

So far then I have emphasised the analogies between probability theory and other mathematical sciences such as Newtonian mechanics. It is now time to point out that these analogies are not perfect, and that there are disanalogies as well. For example, there is an analogy between the application of the falsifying rule (neglecting the probability of a result in the tails of the distribution in comparison with the probability of a result in the distribution's head), and the way in which Kepler's third law was derived from Newton's theory by neglecting the mass of a planet in comparison with that of the Sun. However, this analogy is far from perfect. In the Newtonian case, specific masses (of the Sun, of Mars, etc.) are considered, and a judgement is made about their relative magnitudes. In a different application, different masses would be considered, and different approximations made. The falsifying rule is, by contrast, something which has to be applied in a uniform way, whenever probability hypotheses are compared with frequency data. Thus, I am partly in agreement with an insightful passage of De Finetti's in which he argues that there is a difference between probability theory and other physical sciences. The passage (already quoted on p. 103) runs as follows:

> It is often thought that these objections may be escaped by observing that the impossibility of making the relations between probabilities and frequencies precise is analogous to the practical impossibility that is encountered in all the experimental sciences of relating exactly the abstract notions of the theory and the empirical realities. The analogy is, in my view, illusory: in the other sciences one has a theory which asserts and predicts with certainty and exactitude what would happen if the theory were completely exact; in the calculus of probability it is the theory itself which obliges us to admit the possibility of all frequencies. In the other sciences the uncertainty flows indeed from the imperfect connection between the theory and the facts; in our case, on the contrary, it does not have its origin in this link, but in the body of the theory itself ...
>
> (De Finetti 1937:117)

The falsifying rule for probability statements (FRPS) is not then a specific assumption needed for a particular application, but a general assumption needed for all applications. We can look at the matter in this way. Suppose a non-statistical mathematical science is based on a set of axioms. These axioms are justified if from them we can derive a mass of results which are in agreement with observation. In the case of probability theory, however, we have to adopt a set of axioms *and* a falsifying rule. It is from this system (Σ say) as a whole that we can derive results, such as the empirical laws of probability, which are in agreement with observation. Thus Σ as a whole, including the falsifying rule, is justified by its empirical and practical successes.

But here another problem appears. Suppose the significance level for our FRPS is set at *k* per cent. Then from ? we can infer that the FRPS will lead to a wrong falsification of a true statistical hypothesis in approximately *k* per cent of the cases where it is applied to such hypotheses. In other words, if we are right to rely on the FRPS, we are right to believe that it will give us the wrong answer in *k* per cent of cases of a certain type. We cannot therefore consistently adopt the FRPS. The rule will inevitably lead to inconsistency. We can sum up the situation by saying that the FRPS is practically and empirically successful and yet inconsistent.

This inconsistency should warn us to take care and to handle our falsifying rule judiciously, but it is not in my view fatal to the whole approach. After all, whenever we apply mathematical theories to real situations, there are always numerous possible sources of error, and the inconsistency in the FRPS only adds one more to this number. This additional source of error is something which we can learn to live with, as statistical practice shows.

The Kolmogorov axioms and the propensity theory*

We have already considered the Kolmogorov axioms in the context of the subjective and frequency interpretations of probability. Both these interpretations gave an explicit definition of probability, and it was necessary to prove from this definition that the Kolmogorov axioms held. This we were able to do, apart from the question of countable additivity within the frequency theory. The justification of the Kolmogorov axioms in the context of the propensity interpretation of probability is rather different. The propensity theory (in the version given here) does not offer an explicit definition of probability from which the axioms can be derived. It rather regards probability as implicitly characterised by a set of axioms which are designed to provide a mathematical theory of observed random phenomena. The axioms are justified by showing that from them results can be derived which are in agreement with observation. In particular, the Kolmogorov axioms would be justified by showing that we can derive from them the two empirical laws of probability – the Law of Stability of Statistical Frequencies and the Law of Randomness. We have already carried out this derivation in the simple special case of a biased coin. Let us now see what it looks like when the Kolmogorov axioms are considered in their full generality.

The Kolmogorov axioms are normally stated in terms of the concept of *probability space,* which is defined as an ordered triple (Ω, F, P), where Ω is the *sample space* or *attribute space,* F is a *Borel field* of subsets of Ω and P is a real valued function defined on F. The Kolmogorov axioms can now be summarised as a single axiom (Axiom I) which can be stated as follows:

Axiom I (Kolmogorov's axioms)

P is a non-negative, countably additive set function on F such that P(Ω) = 1.

To this axiom there needs to be added a definition of conditional probability. Alternatively, conditional probability could be taken as a primitive notion and characterised by another axiom. To connect these axioms to the world of observation, we must of course add a falsifying rule (FRPS), but it turns out that this is not enough. From the Kolmogorov axioms together with an FRPS, we cannot in fact derive the two empirical laws of probability. Something else is needed. I will next argue that what is needed is another axiom – an Axiom II to be added to the above Axiom I. This Axiom II will be called the *Axiom of Independent Repetitions.* At first sight this may seem a rather odd proposal, but I will show that this extra axiom is nothing other than an explicit formalisation of various informal suggestions which Kolmogorov (1933:§2) makes in the section of his monograph in which he discusses the relation of his theory to experimental data.

As we have seen (p. 117), although Kolmogorov (1933: 3) claims in a footnote to be using the work of Von Mises, his approach is in fact closer to that of the propensity theory. Thus he relates probabilities to the outcomes of sets of repeatable conditions (**S**) rather than to collectives (**C**) of the Von Mises' type. The propensity theory claims that some sets of repeatable conditions have a propensity to produce in a long sequence of repetitions frequencies which are approximately equal to the probabilities. Kolmogorov gives what really amounts to a formulation of this basic principle of the propensity theory. He says:

> One can be practically certain that if the complex of conditions **S** is repeated a large number of times, *n,* then if *m* be the number of occurrences of event A, the ratio *m*/*n* will differ very slightly from P(A).
>
> (1933:4)

This principle relates probabilities P(A) to the frequencies obtained by repeating the underlying conditions **S** a large number of times. It is not really possible to formalise such a principle within the usual framework of probability spaces, because these contain no mention of the repeatable conditions **S**. To overcome this difficulty, I suggest that we introduce the concept of a *probability system,* defined as an ordered quadruple (**S**, Ω, F, P), where (Ω, F, P) is a probability space in the ordinary sense given above, and Ω is the set of possible outcomes of the repeatable conditions **S**. I will take it as a basic premise of the propensity theory that if (**S**, Ω, F, P) is any probability system, and A ε F is any event, then, if the conditions **S** are repeated a large number *n* of times and A occurs *m*(A) times, it is practically certain that

$$\frac{m(\mathrm{A})}{\mathrm{n}} \approx \mathrm{P}(\mathrm{A}) \qquad (7.3)$$

The meaning of 'practically certain' and '≈' will of course be made more precise in due course using the falsifying rule. However, I take Equation 7.3 as an informal principle which lies at the heart of the propensity theory. It was advocated by Kolmogorov and in Popper's earlier version of the propensity theory, and indeed commends itself strongly to common sense. At any rate I will assume Equation 7.3 in what follows. It turns out, perhaps surprisingly, that it has some important consequences.

Before we can draw these consequences, it will be necessary to analyse the concept of repeatability in more detail than we have done hitherto. Let us begin with the obvious point that any two alleged repetitions will be found on closer inspection to differ in many respects. Consider for example two tosses of a coin which would ordinarily be regarded as repetitions. Closer inspection might reveal that in one case the head had been uppermost before the toss was made and in the other the tail. Moreover, even if every macroscopic property of the tossing procedure did appear to be the same in the two cases, there would still be the difference that the two tosses occurred at different times. We must therefore regard two events as repetitions not if they are the same in every way (which is impossible), but if they are the same in a well-specified set of ways. Two events are not in themselves repetitions. The question of whether they are such depends on how we are proposing to describe them. Any two events, however similar, will differ in some respects and this could bar them from being thought of as repetitions. Conversely, any two events, however dissimilar, will agree in some respects and this could lead to their being considered as repetitions. In fact, a sequence is a sequence of repetitions only relative to some set of common properties or conditions. These considerations suggest the following definition. A sequence is a sequence of repetitions relative to a set **S** of conditions or properties if each member of the sequence satisfies every condition of **S**, and *irrespective of how the members differ in other respects.* This definition is all right as far as it goes, but it will be convenient to modify it in the light of another aspect of the matter.

In any sequence of repetitions there is characteristically not only a set of constant features but also some variable feature. Usually this variable parameter is time, as in the case of a sequence of tosses of a coin, but this need not be so. Consider, for example, twenty students carrying out the 'same' experiment at the same time. Here the variable parameter is spatial position. Yet again the variable parameter may include both spatial and temporal constraints. This leads to the following difficulty. Consider the case of the twenty students and suppose that they are performing an electrical experiment. We might not want to consider such experiments as repetitions even though some set of defining conditions were satisfied if, in addition, the pieces of apparatus were so close together that some kind of magnetic interference occurred. We would require in effect that the experiments should be sufficiently widely spaced in position. Similarly, in the temporal case we might want the various events to be sufficiently widely spaced in time. Of course, we could regard this matter as being included in the relevant set of conditions **S**, but I think it will be better to treat it separately by introducing the concept of a *spacing condition.* We will henceforth require that any sequence of repetitions must involve a spacing condition s stating that the elements of the sequence must be separated in such and such a way relative to some variable parameter (e.g. time or spatial position or a combination of both).

We can illustrate the value of the concept of spacing condition by using it to analyse an interesting example of Popper's. Popper is considering the probability of someone of a particular age surviving another year or twenty more years, and argues that this is not a function just of that person's state of health. As he says:

> Nevertheless, the view that the propensity to survive is a property of the state of health *and not* of the situation can easily be shown to be a serious mistake. As a matter of course, the state of health is very important – an important aspect of the situation. But as anybody may fall ill or become involved in an accident, the progress of medical science – say, the invention of powerful new drugs (like antibiotics) – changes the *prospects* of everybody to survive, whether or not he or she actually gets into the position of having to take any such drug. The *situation* changes the possibilities, and thereby the propensities.
>
> (Popper 1990:14–15)

We can deal with Popper's point here by saying that when we are considering the probability of surviving for a particular number of years, we should take our set of repeatable conditions **S** as defining a specific state of health, and the spacing condition s as stating that repetitions should consist of individuals of the age in question in that state of health at a particular time, and not of such individuals in the state of health at different times. This gives a satisfactory account of Popper's example without having to assign propensities to unrepeatable states of the universe.

We can now reformulate our definition of 'sequence of repetitions' as follows. A sequence of events is a sequence of repetitions relative to a set of conditions $\mathbf{S}_s$ which include a spacing condition s, provided that all the conditions **S** are satisfied by each event, and the events are spaced as required by s. A set $\mathbf{S}_s$ of conditions is repeatable provided that an indefinitely long sequence of repetitions relative to $\mathbf{S}_s$ is in principle possible.

The point of giving the above analysis is that it enables us to raise and answer the following important question: does repeatability imply independence? In fact it is easy to show that repeatability does not imply independence, since we can give examples of sets of repeatable conditions whose outcomes are dependent. Indeed almost any example of a Markov chain would serve as an example of this kind. We can therefore consider again two examples of Markov chains given above (p. 78). These were (a) the game of red or blue and (b) sequences of dry or rainy days during the rainy season in Tel Aviv.

In the case of the game of red or blue, the set of conditions **S** specifies that a fair coin be tossed, that one be added to the score if the result is heads and subtracted if the result is tails and that if the resulting score is greater than or equal to 0, the result be given as blue, whereas if it is less than 0, the result be given as red. The spacing condition s is that successive goes of the game be considered. This set of repeatable conditions $\mathbf{S}_s$ accords with the analysis just given, and yet the outcomes in a sequence of repetitions are clearly dependent. In the other example, the set of conditions **S** specifies that we observe whether in Tel Aviv there is any rain on a particular day and record the result as rainy if some rain does fall, and otherwise as dry. The spacing condition s specifies that we consider successive days during the rainy season of December, January and February. The set of conditions $\mathbf{S}_s$ is again a bona fide set of repeatable conditions, but the outcomes are dependent. It is more probable that a particular day will be dry if the previous day was dry (probability = 0.75) than if the previous day was wet (probability = 0.34).

Granted then that repeatability does not imply independence, at least two different courses of action become possible. The first of these would be to introduce probabilities for the outcomes of sets of repeatable conditions $\mathbf{S}_s$ regardless of whether the repetitions of these conditions were independent or not. A second, alternative, course of action would be to ascribe probabilities only to the outcomes of sets of repeatable conditions $\mathbf{S}_s$ whose repetitions are independent. I will now argues in favour of the second of these two courses of action.

Suppose we did adopt the first course of action and were prepared to ascribe probabilities to the outcomes of sets of repeatable conditions $\mathbf{S}_s$, even in cases where the repetitions of these conditions were dependent. In particular, we would be prepared to introduce the probabilities Prob(red | $\mathbf{S}_s$) and Prob(blue | $\mathbf{S}_s$), where $\mathbf{S}_s$ are the sets of conditions defining the game of red or blue. Suppose we set up the game so that red or blue are exactly symmetrical, then presumably we would have in this case Prob(red | $\mathbf{S}_s$) = Prob(blue | $\mathbf{S}_s$) = $^1/_2$. But now let us recall the curious features of the game of red or blue quoted earlier (p. 78) from Feller (1950:82–3). Feller shows that if the game is played once a second for a year, i.e. 31,536,000 repetitions, there is a probability of 70 per cent that the more frequent colour will appear for a total of 265.35 days, or about 73 per cent of the time, whereas the less frequent colour will appear for only 99.65 days, or about 27 per cent of the time. This shows clearly that Equation 7.3 does not hold for the game of red or blue. This equation states that for a large number n of repetitions, it is practically certain that the observed relative frequency of an attribute A [m(A)/n] will be approximately equal to its probability [P(A)]. Now 31,536,000 is surely a very large number of repetitions, and yet in 70 per cent of the cases in which this number of repetitions of the game of red or blue were carried out, the observed frequency of each attribute would differ from its probability (0.5) by 0.23, i.e. almost half the largest divergence possible.

I have already argued that Equation 7.3 is the core of the propensity theory, and intuitively highly plausible. If, however, we allow the ascription of probabilities to the outcomes of repeatable conditions whose repetitions are dependent, then Equation 7.3 may well fail. I suggest therefore that we should not assign probabilities in the case of repeatable conditions whose repetitions are dependent. Indeed it would in my view be highly dangerous to develop a theory along these lines. Since it is almost automatic to identify probabilities approximately with frequencies in a long sequence of repetitions, a theory in which such an identification was sometimes completely wrong would be very liable to mislead. I therefore conclude that we should assign probabilities only to the outcomes of sets of repeatable conditions whose repetitions are independent. This amounts to a new postulate which I will call the *Axiom of Independent Repetitions.* It can be formulated as follows.

Consider a sequence of repetitions of $\mathbf{S}_s$. Suppose we select a particular n-tuple of these repetitions say (i_1, i_2, ..., i_n). This procedure can be repeated over and over. In each case we form a sequence of repetitions of $\mathbf{S}_s$, and then select the same n-tuple of these repetitions. The procedure is thus itself a set of repeatable conditions which we will denote by $\mathbf{S}_s{}^n$. Suppose now we start with a probability system ($\mathbf{S}_s$, Ω, F, P). Let Ω^n denote the n-fold Cartesian product of Ω. Let F^n be a Borel field of subsets of Ω^n defined as follows: we consider the set (S say) of all Cartesian products $A_1 \times A_2 \times \ldots \times A_n$, where each A_i ε F and take F^n to be the minimum Borel field containing F. We have already summarised the Kolmogorov axioms as Axiom I, and we can now state our new axiom as Axiom II.

Axiom II (Axiom of Independent Repetitions)

If ($\mathbf{S}_s$, Ω, F, P) is a probability system, so is ($\mathbf{S}_s{}^n$, Ω^n, F^n, $P^{(n)}$) for any n, where the measure $P^{(n)}$ on F^n is the n-fold product measure of the measure P on F.

From Axioms I and II and of course the falsifying rule for probability statements, the two empirical laws of probability can be derived. The derivation is just the same as in the case of the biased coin, where, corresponding to our Axiom of Independent Repetitions, we assumed that the tosses were independent. The derivation of these empirical laws, which are confirmed by a mass of data, justifies the adoption of Axioms I and II, and indeed the FRPS. This is the justification of the Kolmogorov axioms (Axiom I) within the propensity theory.

Note that there is no difficulty here about countable additivity, which raised so many problems in some of the other interpretations of probability. The axioms are being set up in order to explain observations of random phenomena. We want the simplest mathematical theory which will explain the observations, and, since countable additivity is simpler from a mathematical point of view than finite additivity and does the explanatory work very satisfactorily, we are quite justified in adopting it. So countable additivity is justified by the propensity theory, whereas Von Mises was forced simply to postulate countable additivity without being able to offer an empirical justification for this extra axiom (see p. 110–11). This is a point in favour of the propensity theory since countable additivity is so much more convenient than finite additivity, and it is in fact used in practice by nearly all mathematicians who work in probability theory.

I will conclude this section and the chapter by making a number of comments on the Axiom of Independent Repetitions. First of all, it is necessary to answer an objection which might be made to the axiom. The axiom, it might be argued, limits probability theory to the special case of independence, whereas probability theory deals with dependent events, e.g. Markov chains, as well as independent ones. This argument is, however, mistaken, since it is perfectly possible – indeed straightforward – to deal with Markov chains and other cases of dependent events in a framework which includes the Axiom of Independent Repetitions. We can show this by considering again our two examples of Markov chains. The general idea is that we deal with Markov chains by taking a complete sequence of results of the chain as a single point in the attribute space Ω, so that the repeatable conditions are those that define the chain as a whole, and the independent repetitions are independent realisations of the entire chain. Thus in the game of red or blue, our repetitions are different games all starting from the same initial position. These different games are quite independent, even though the goes which compose each game are highly dependent. So the Axiom of Independent Repetitions is satisfied. Similarly in the Tel Aviv example, repetitions might be observed sequences of dry or rainy days during the rainy season in successive years. These sequences could well be independent, even though the outcomes on successive days with in any given sequence are highly dependent. In this way the Axiom of Independent Repetitions would be satisfied.

This answers an objection which might be made to the Axiom of Independent Repetitions. Let us now consider some points in its favour. To begin with it enables us to solve a problem which arose when we examined the Kolmogorov axioms in the context of Von Mises' theory (p. 112). All the Kolmogorov axioms seemed to correspond to just the first of Von Mises' two axioms (the axiom of convergence), and there was nothing in the Kolmogorov axioms corresponding to the axiom of randomness. The Axiom of Independent Repetitions fills this gap in Kolmogorov's treatment and provides something corresponding to Von Mises' axiom of randomness. Indeed we have tried to emphasise this correspondence by giving a formulation in which there are two axioms (I and II). These correspond to Von Mises' two axioms. The link between the Axiom of Independent Repetitions (Axiom II) and Von Mises' axiom of randomness corresponds to our view (see p. 154 above) that Von Mises' concept of randomness is, within the propensity theory, reduced to that of independence.

Next let us consider another problem which arose within Von Mises' theory, and which was mentioned earlier (p. 90). A mathematical collective consists of an ordered sequence, numbered 1, 2, and so on. However, there are several examples of empirical collectives which are not naturally ordered. For example, the plants in a field, or the molecules in a gas do not occur in a particular sequence. Is it legitimate to represent such unordered empirical collectives by an ordered sequence?

Of course, in the propensity theory we have replaced collectives by sets of repeatable conditions $\mathbf{S}_s$. However, essentially the same problem still arises, for it is assumed that on repeating $\mathbf{S}_s$ we get an ordered sequence: a first repetition, a second repetition, and so on. But what then about examples in which there is no natural order? In the present framework these examples occur where the spacing parameter s in $\mathbf{S}_s$ is literally to do with spatial distances. Consider the example of the molecules of a gas at a particular time *t*. Our repeatable conditions specify that we must select a particular molecule (the outcome might be its instantaneous velocity at *t*). Repetitions of the conditions are obtained by taking different, i.e. spatially distinct, molecules. Now evidently there is no natural ordering of the molecules and we must impose one arbitrarily to get our ordered sequence of repetitions. This might seem a dubious procedure but provided the Axiom of Independent Repetitions holds it is easily shown to be legitimate. Since the observations are independent it does not matter what order we take them in. If it is convenient mathematically to impose a particular order, we are quite entitled to do so.

Kolmogorov makes the following very interesting observation (I have altered his notation at one point to agree with ours):

> ... the theory of probability can be regarded from the mathematical point of view as a special application of the general theory of additive set functions. One naturally asks, how did it happen that the theory of probability developed into a large individual science possessing its own methods?
>
> In order to answer this question, we must point out the specialisation undergone by general problems in the theory of additive set functions when they are proposed in the theory of probability.
>
> The fact that our additive set function P(A) is non-negative and satisfies the condition P(O) = 1, does not in itself cause new difficulties. Random variables ... from a mathematical point of view represent merely functions measurable with respect to P(A), while their mathematical expectations are

> abstract Lebesgue integrals. (This analogy was explained fully for the first time in the work of Fréchet.) The mere introduction of the above concepts, therefore, would not be sufficient to produce a basis for the development of a large new theory.
>
> Historically, the independence of experiments and random variables represent the very mathematical concept that has given the theory of probability its peculiar stamp....
>
> We thus see, in the concept of independence, at least the germ of the peculiar type of problem in probability theory.
>
> (1933:8–9)

But if independence is the key concept which differentiates probability theory from other related branches of mathematics, should this concept not appear in the axioms of the theory? In fact it does so if we adopt the Axiom of Independent Repetitions. However, this observation can be carried further. When discussing the subjective theory, we remarked (see p. 75 above) that in a certain sense the concept of exchangeability is the equivalent within the subjective theory of the objectivist's notion of independence. Although we can define independence within the subjective theory by the same formulas which are used in the objective approach, it turns out that, within the subjective theory, the assumption of independence is equivalent to the assumption that no learning from experience can occur. So the assumption of independence will rarely, if ever, be made within the subjective theory. Where an objectivist assumes independence, a subjectivist will assume exchangeability. So independence is not characteristic of probability theory in general, but of the objective interpretation of probability. This suggests that the Axiom of Independent Repetitions might serve to differentiate the objective interpretation of probability from the subjective. I will next argue that this is indeed the case.

It follows from our earlier discussions that the Kolmogorov axioms can be interpreted either objectively or subjectively. It is worth noting, however, that, in either of these interpretations, the conditional probabilities as written explicitly in the system are actually abbreviated in a significant fashion. To see this let us begin with the objective case. Here the sample space (or attribute space) Ω is the set of possible outcomes of some repeatable conditions $\mathbf{S}_s$. Within Kolmogorov's formalism we write conditional probabilities in the form $P(A \mid B)$ where A, B are subsets of Ω. As already pointed out, however, $P(A \mid B)$ is really an abbreviation for $P(A \mid B \,\&\, \mathbf{S}_s)$, although the underlying repeatable conditions $\mathbf{S}_s$ are never written out explicitly within Kolmogorov's formalism.

Exactly the same applies in the subjective interpretation. Let e and f be two propositions stating the occurrence of particular events E and F. Then we only write explicitly conditional probabilities of the form $P(e \mid f)$, but here $P(e \mid f)$ is really an abbreviation for $P(e \mid f \,\&\, K)$, where K is the background knowledge assumed by the individual giving the subjective probability.

The key point to note is this. In both cases conditional probabilities are written in an abbreviated form, but what must be added to expand the abbreviation is different in the two cases. In the objective case, it is a set of repeatable conditions $\mathbf{S}_s$, while in the subjective case it is the body of background knowledge K assumed by the individual in question. So, if we make this expansion explicit, we differentiate between the two interpretations. But this is just what we have done by moving from Kolmogorov's concept of probability space (Ω, F, P) to the concept of probability system ($\mathbf{S}_s$, Ω, F, P) and formulating explicitly the Axiom of Independent Repetitions.

This sheds some new light on the significance of Kolmogorov's axioms and their central rôle in probability theory. The axioms are sufficiently abstract to be satisfied by both the subjective and propensity interpretations. They thus exhibit the mathematical or structural features in common to these interpretations. If, however, we want to differentiate the objective from the subjective interpretation, we can do so by adding another axiom – the Axiom of Independent Repetitions. If we want to justify the resulting system by relating it to observation, then we have to add a falsifying rule. In this way a bridge is created from the abstract mathematical axioms to the world of experience.

8 Intersubjective probability and pluralist views of probability

The discussions of the preceding chapters have led to what may seem to be an excessively sharp polarisation between the subjective view, in which probability is the degree of belief of an individual, and the objective view, in which probabilities are features of the material world like charges or masses. In this chapter I want to try to moderate this difference by suggesting that there are some intermediate cases. Accordingly in the section ‘Intersubjective probability’, I will introduce a further interpretation of probability – the *intersubjective* – which, as the name suggests, lies somewhere between the subjective and the objective. Then in the section ‘The spectrum from subjective to objective’, I will try to show that there is a spectrum of positions between the subjective and the fully objective, and I will try to analyse the character of this spectrum. This analysis naturally suggests that there is not a single notion of probability, but rather several different, though interconnected, notions of probability which apply in different contexts. In the section ‘Pluralist views of probability’, I will discuss such pluralist views of probability.

Intersubjective probability

The starting point of the subjective theory of probability was the degree of belief of a particular individual whom we called Mr B. We imagined a psychologist Ms A, who undertook to measure Mr B’s degree of belief by getting him to bet in a carefully specified betting situation. So the theory is concerned with degrees of belief of particular individuals. However, this abstracts from the fact that many, if not most, of our beliefs are social in character. They are held in common by nearly all members of a social group, and a particular individual usually acquires them through social interactions with this group. If we accept Kuhn’s (1962) analysis then this applies to many of the beliefs of scientists. According to Kuhn, the scientific experts working in a particular area, nearly all accept a paradigm which contains a set of theories and factual propositions. These theories and propositions are thus believed by nearly all the members of this group of scientific experts. A new recruit to the group is trained to know and accept the paradigm as a condition for entry to the group. Much the same considerations apply to other social groups such as religious sects, political parties, and so on. These groups have common beliefs which an individual usually acquires through joining the group. It is actually quite difficult for individuals to resist accepting the dominant beliefs of a group of which they form part, though of course dissidents and heretics do occur. One striking instance of this is that individuals kidnapped by a terrorist organisation do sometimes, like Patty Hearst, adopt the terrorists' beliefs. All this seems to indicate that as well as the specific beliefs of a particular individual, there are the consensus beliefs of social groups. Indeed the latter may be more fundamental than the former. In Chapter 4 subjective probabilities were introduced by using the Dutch book argument. What I want to show now is that we can extend the Dutch book argument to social groups, and this extension will introduce the concept of what I will call *intersubjective probability.*

Let us begin by recalling the definition of betting quotient given previously (p. 55). We imagined that Ms A (a psychologist) wanted to measure the degree of belief of Mr B in some event E. To do so, she gets Mr B to agree to bet with her on E under the following conditions. Mr B has to choose a number q (called his betting quotient on E), and then Ms A chooses the stake S. Mr B pays Ms A qS in exchange for S if E occurs. S can be positive or negative, but $|S|$ must be small in relation to Mr B's wealth. Under these circumstances q is taken to be a measure of Mr B's degree of belief in E.

In order to extend this to social groups, we can retain our psychologist Ms A, but we should replace Mr B by a set $\mathbf{B} = (B_1, B_2, ..., B_n)$ of individuals. For simplicity, let us take $n = 2$ initially. We then have the following theorem.

Theorem 1

Suppose Ms A is betting against $\mathbf{B} = (B_1, B_2)$ on event E. Suppose B_1 chooses betting quotient q_1 and B_2 q_2. Ms A will be able to choose stakes so that she gains money from **B** whatever happens *unless* $q_1 = q_2$.

Proof

Suppose without loss of generality that $q_1 > q_2$. Suppose Ms A chooses $S > 0$ on her bet with B_1, and $-S$ on her bet with B_2. Then if E occurs, Ms A's gain G_1 is given by:

$$G_1 = q_1S - S - q_2S + S = (q_1 - q_2)S$$

If E does not occur, Ms A's gain G_2 is given by:

$$G_2 = q_1S - q_2S = (q_1 - q_2)S$$

It is clear that $G_1 > 0$ and $G_2 > 0$, unless $q_1 = q_2$.

Acknowledgement

Theorem 1 was suggested to me by Ryder (1981). In this important paper, Ryder (1981:165) gives a result which is a special case of Theorem 1. Ryder uses this result to draw philosophical conclusions which are different from my own. I will use Theorems 1, 2 and 3 to introduce the concept of intersubjective probability. However, I see intersubjective probabilities as additional to, rather than in contradiction with, subjective probabilities. Ryder, on the other hand, regards his result as showing that the whole subjective approach to probability based on the Dutch book argument is not viable. I will state and discuss Ryder's argument on this point in a moment.

The generalisation from 2 to n is perfectly straightforward.

Theorem 2
Suppose Ms A is betting against $\mathbf{B} = (B_1, B_2, ..., B_n)$ on event E. Suppose B_i chooses betting quotient q_i. Ms A will be able to choose stakes so that she gains money from **B** whatever happens *unless* $q_1 = q_2, = ... = q_n$.

Proof [1]
Suppose the q_i are not all equal. Then there must exist q_j and q_k such that $q_j > q_k$. Suppose Ms A chooses $S > 0$ on her bet with B_j, $-S$ on her bet with B_k, and $S = 0$ on her bet with B_i where $i \neq j$ and $i \neq k$. Then arguing as in the proof of Theorem 1, we conclude that Ms A gains money from **B** whatever happens. Thus Ms A can gain money from B whatever happens, unless $q_1 = q_2 = ... = q_n$.

Theorem 3
Suppose Ms A is betting against $\mathbf{B} = (B_1, B_2, ..., B_n)$ on events $E_1, ..., E_r$ where $r \geq 1$. Suppose B_i chooses betting quotient q_{ij} on event E_j. If (a) $q_{1j} = q_{2j} = ... = q_{nj} = q_j$ for $1 \leq j \leq r$, and (b) the q_j satisfy the standard axioms of probability, then it will not be possible for Ms A to make a Dutch book against **B.**

Proof
If condition (a) is satisfied, then the group can be considered as a single individual with betting quotient q_j on E_j for $1 \leq j \leq r$. The result then follows from the converse of the Ramsey–De Finetti theorem using condition (b).

Informally what Theorems 1, 2 and 3 show is this. Let **B** be some social group. Then it is in the interest of **B** as a whole if its members agree, perhaps as a result of rational discussion, on a common betting quotient rather than each member of the group choosing his or her own betting quotient. If a group does in fact agree on a common betting quotient, this will be called the *intersubjective* or *consensus* probability of the social group. This type of probability can then be contrasted with the *subjective* or *personal* probability of a particular individual.

The Dutch book argument used to introduce intersubjective probability shows that if the group agrees on a common betting quotient, this protects them against a cunning opponent betting against them. This then is a particular mathematical case of an old piece of folk wisdom, the claim, namely, that solidarity within a group protects it against an outside enemy. This point of view is expressed in many traditional maxims and stories. A recent example occurs in Kurosawa's film *Seven Samurai.* In one particular scene Kambei the leader of the samurai is urging the villagers to act together to repel the coming attack by bandits. 'This is a rule of war.' he says 'Collective defence protects the individual. Individual defence destroys the individual.'

Returning, now, to our main theme, the question arises: 'under what conditions will a social group form an intersubjective probability?' The following conditions seem to be of crucial importance:

1 Common Interest: The members of the group must be linked by a common purpose; whether the common purpose leads to solidarity or rivalry within the group does not matter much; the important point is that the members have an interest in acting together and reaching consensus; love or fear would create, in this case, similar bonds. The common purpose might be financial, but need not be; for example, a group of soldiers might have the common purpose of taking an enemy position with the minimum injury and loss of life to the group.
2 Flow of Information: There must be a flow of information and exchange of ideas between members, though it does not matter whether the communication is organised centrally or peripherally or whether it is direct (between any two members) or indirect (through the intervention of third parties).

I will next make a few comments on these two conditions. Condition 1 (Common Interest) means that the size and composition of the group can change since individual members may decide to break away if and when they reckon that they can gain by 'going it alone'; equally, new members may join the group when they recognise a community of purpose with it. The common purpose must be strong enough to bond the members together in relation to particular events, though this need not rule out individual members planning to break away or to gain at the expense of others on different issues. Another related point is that the propositions whose intersubjective probabilities are sought must be connected with the common purpose. Consider, for example, the group of Italian expatriates living in London. This group might well form a consensus probability regarding the question of whether there will be new regulations within a few years making it possible for Italian nationals living in the UK to vote in local elections. However, it seems unreasonable that the group should form an intersubjective probability concerning the number of king penguins on Elephant Island in the South Pacific.

Condition 2 (Flow of Information) has implications regarding conditional probabilities. In the subjective theory we write the probability which Mr B assigns to event E as P(E), but this probability is really conditional, and, if written explicitly, should appear as $P(E \mid K)$, where K is the set of beliefs which constitute Mr B's assumed background knowledge. Intersubjective probabilities are also of the form $P(E \mid K)$, but K is now the background knowledge of the group. This may be more extensive than the knowledge possessed by any individual members of the group. Since there is flow of information and exchange of ideas within the group, if one member has a piece of relevant knowledge which the others lack, he or she can communicate it to the others. Similarly the logical and mathematical powers of a group will normally exceed those of any of its members. If anyone makes a logical mistake, this is characteristically exposed and corrected by someone else, and even the finest mathematicians make the occasional logical blunder.

There may be a problem in satisfying conditions 1 and 2 if the group is very large. However, consensus probabilities may still be possible in this case provided there is an agency or association or union to organise the group including the flow of information within it.

The concept of intersubjective probability has, so I believe, possible applications in a number of different areas: one of these is economics, and another is the confirmation of scientific hypotheses.[2] I do not however want to suggest that intersubjective probabilities should completely replace subjective probabilities. The use of the first concept does not exclude the use of the second, but rather demands its use. If, for example, P(E) is the intersubjective probability assigned to E by the social group $\mathbf{B} = (B_1, B_2, ..., B_n)$, then each member B_i of **B** assigns the subjective probability P(E) to E. Moreover there may well be sets of individuals who do not reach a consensus and who therefore have each a subjective probability without there being any intersubjective probability. Then again a social group may reach a consensus which is accepted by nearly all its members, while containing one or two dissidents who have subjective probabilities which differ from the intersubjective probability of the group. In my opinion these various possibilities show that both subjective and intersubjective probabilities are needed for the analysis of human belief.

Having just defended the concept of subjective probability, it is now appropriate to consider Ryder's objection to subjectivism which was mentioned earlier in this section. Ryder states his argument as follows:

> Subjectivists accept that different individuals have different degrees of belief, but not much thought has been given to applying the Dutch Book argument to the situation where there is more than one person.
>
> If we have two (or more) people with different degrees of belief in the same simple event E, a Dutch Book can be made against them. This is just as 'disastrous' and 'obviously unreasonable' as it is for an individual. It means that Subjectivists never actually make the bets which are envisaged by the Dutch Book argument. If they did someone could come along and find two or more subjectivists with different degrees of belief and make a system of bets which would result in a certain loss to the subjectivists considered as a group.
>
> (1981:165)

This argument of Ryder's is both plausible and ingenious, but I think that it can, nonetheless, be answered. To do so let us consider our set of individuals $\mathbf{B} = (B_1, B_2, ..., B_n)$ and the experimental psychologist Ms A. Suppose first that Ms A makes a Dutch book against B_1. It is very likely that B_1 will regard this as 'disastrous' and 'obviously unreasonable', since he or she will lose money whatever happens. Suppose, however, that Ms A makes a Dutch book against the set **B** as a whole, but *not* against B_1 in particular. Will B_1 regard this situation as 'disastrous' and 'obviously unreasonable'? The answer is that B_1 may do so, but he or she need not necessarily do so. To see this let us consider two different, indeed extreme, cases.

1 B_1, B_2, ..., B_n have formed an arrangement whereby any gains or losses they make individually in their various economic activities are pooled, and the total divided equally between the individual members of the group. In this case, if Ms A makes a Dutch book against the set as a whole, then each of its members (including B_1) will suffer. Thus B_1 has to regard this situation as 'disastrous' and 'obviously unreasonable'. Note that this example has been constructed so that our condition 1 (Common Interest) is satisfied. Given such an arrangement, it would obviously be desirable for the group to ensure that condition 2 (Flow of Information) is also satisfied, so that the group can form a consensus through discussion and have an intersubjective probability, thus rendering it impossible for Ms A to make a Dutch book against them.
2 B_2, ..., B_n are more or less randomly selected individuals whom B_1 neither knows nor cares about. In this case Ms A is likely to be able to make a Dutch book against the group **B** as a whole, but B_1 is unlikely to regard this as 'disastrous' and 'obviously unreasonable'. Provided no Dutch book can be made against B_1 personally, why should he or she care about what happens to the other unknown members of the group?

The key point is that the extension of the Dutch book argument to groups is only significant for groups which have a common interest. The argument shows that such groups ought to establish communication and flow of information within the group so that they can form through discussion a consensus or intersubjective probability. Only in this way can the group as a whole protect itself against cunning opponents. It is a matter of common experience that there do exist such groups with a common interest and that they do often reach consensus in their beliefs.

If, however, we are dealing with a group which lacks a common interest, the extension of the Dutch book argument to groups has no validity, for each individual will then be indifferent to what happens to the other members of the group. In this case each individual will form his or her own subjective probability without any regard for the beliefs of the others.

One helpful way of regarding the intersubjective interpretation of probability is to see it as intermediate between the logical interpretation of the early Keynes and the subjective interpretation of his critic Ramsey. According to the early Keynes, there exists a single rational degree of belief in some conclusion c given evidence e. If this were really so, we would expect nearly all human beings to have this single rational degree of belief in c given e, since, after all, most human beings are rational. Such a broad consensus does indeed exist as regards deductive logic. Nearly all human beings who have acquired the technical background needed to understand the question will agree on whether a given chain of logical deductions is valid or invalid. Of course, this consensus is not complete. There are indeed a few intuitionists and other believers in various forms of non-standard logic. However, the agreement in judgement is considerable even if not total.

Far otherwise is the case of judging the degree of belief which evidence e warrants in conclusion c in situations in which e does not logically entail c. Here different individuals may come to quite different conclusions even though they have the same background knowledge and expertise in the relevant area, and even though they are all quite rational. A single rational degree of belief on which all rational human beings should agree seems to be a myth.

So much for the logical interpretation of probability, but the subjective view of probability does not seem to be entirely satisfactory either. Degree of belief is not an entirely personal or individual matter. We very often find that an individual human being belonging to a group which shares a common outlook has some degree of common interest and is able to reach a consensus as regards its beliefs. Obvious examples of such groups would be religious sects, political parties or schools of thought regarding various scientific questions. For such groups the concept of intersubjective probability seems to be the appropriate one. These groups may be small or large, but usually they fall short of embracing the whole of humanity. The intersubjective probability of such a group is thus intermediate between a degree of rational belief (the early Keynes) and a degree of subjective belief (Ramsey).

When Keynes propounded his logical theory of probability, he was a member of an elite group of logically minded Cambridge intellectuals (the Apostles). In these circumstances, what he regarded as a single rational degree of belief valid for the whole of humanity may have been no more than the consensus belief of the Apostles. However admirable the Apostles, their consensus beliefs were very far from being shared by the rest of humanity. This became obvious in the 1930s when the Apostles developed a consensus belief in Soviet communism, a belief which was certainly not shared by everyone else.

The spectrum from subjective to objective

In the last section I showed how, starting with subjective probability and its foundation in the Dutch book argument, we could move in the direction of greater objectivity by introducing intersubjective probabilities. I will now try to show that we can divide objective interpretations into those which are fully objective, and those which involve some subjective (or human) element. This will enable us at the end of the section to construct a spectrum stretching from the fully subjective to the fully objective.

I will use the phrase *fully objective* to refer to things which are completely independent of human beings. An obvious example of such a thing is the Sun. This produced and emitted energy in the time of dinosaurs before any human beings existed, and it would continue to produce and emit energy in just the same way if all human beings were vaporised tomorrow. In fact, human activity has so far had no effect on the Sun. It is thus fully objective. The case of the Sun can be contrasted with that of a cup. Now a cup is a material object and therefore in some sense objective, but it is obviously not human independent. It was made by human beings for human purposes. Indeed we could say that if all human beings were vaporised tomorrow without other objects being affected, the cup, while remaining a material object, would cease to be a cup. A cup is something used to drink liquids, and, if an object is no longer used in this way, it ceases to be a cup.

I will call something which is objective, but *not* human independent, *artefactual.* Here artefactual is of course intended to cover human material artefacts such as cups, but has a somewhat wider sense. This can be illustrated by the example of the heavenly constellations. Let us consider what is probably the best known constellation – the Plough or Big Dipper. This is a group of stars which is easily recognised in the night sky in the northern hemisphere. We cannot say that the Plough is subjective, because anyone with a little instruction can pick it out from the surrounding stars. Nor can we say that it is merely intersubjective like group belief, since it is composed of stars which certainly exist objectively. On the other hand, we cannot say that is fully objective because there is no real physical connection between the stars of which it is composed. This is illustrated in Figure 8.1, which shows the distances between the stars of the Plough.

As can be seen, the stars which seem to form a natural grouping from the point of view of a human observer are actually at very great distances from each other. For example, the stars marked a and e are about fifty light years apart. In other constellations the distances between the stars are even greater. For example, the stars in the constellation of Centaurus are at distances which vary from four light years to 325 light years. Another indication of the arbitrariness of the constellations is the fact that the star groupings used in the civilisation of ancient China were different from those of Western Europe. Constellations are not what Duhem (1904–5:24–30) would have called a *natural classification*.

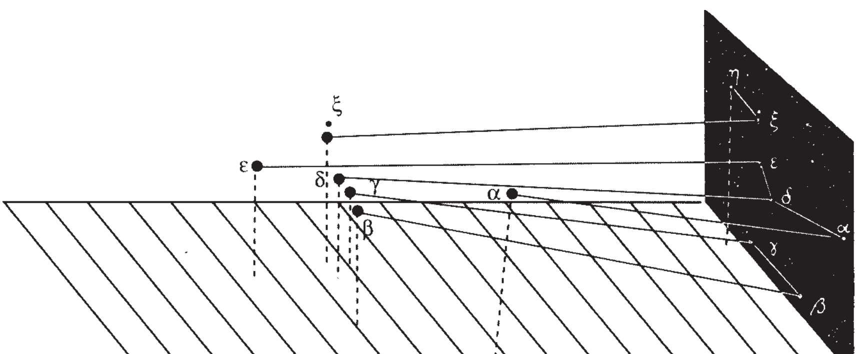


Figure 8.1 Distances in light years between the stars of the Plough (or Big Dipper)

Despite this, however, constellations are in some sense objective, and I propose to classify them like cups as artefactual. The idea is that in both cases there is an underlying material in the natural world. For the cup, this might be clay, and for the constellations, the stars. However, this underlying material is shaped by humans. In the case of the cup, the shaping is a physical process. In the case of the constellations, it is the more intellectual process of choosing to group a set of stars together, and giving that group a particular name. If humans had not existed there would have been no constellations, but equally if the stars had not existed there would have been no constellations. Constellations are a product of the interaction between human beings and the natural world.

Another important point in this connection is that the apparent stability of the shapes of constellations is due to the time scale of human beings.[3] This is illustrated in Figure 8.2, which shows what the Plough would have looked like 100,000 years ago, and what it will look like in 100,000 years time. If we imagine beings for whom 100,000 years was experienced subjectively as we experience a second, the constellations would not have constant shapes but would appear to be continually changing shape. Such beings would therefore not have formed our concept of a constellation as a fixed arrangement of stars. Suppose, conversely, that there were beings who experienced subjectively what for us is a few seconds as hundreds of years. These beings would perceive as fixed and relatively unchanging objects things which for us are entirely transient like the waves of the sea.

A further example of the artefactual is provided by the micro-particles of quantum mechanics, such as electrons or photons. Bohr's resolution of the wave–particle duality was to say that relative to one experimental arrangement an electron is a wave, whereas relative to another it is a particle. Perhaps reality in itself has a holistic character so that it is somewhat arbitrary to divide it into electrons and photons, just as it is arbitrary to divide the stars in the sky into constellations. Or again just as a potter can use one mould to turn the raw clay into a cup and another to turn it into a plate, so the physicist can use one experimental arrangement to turn the electron into a wave and another to turn it into a particle. This does not mean that the electron lacks an objective existence. Electrons have just as much of an objective existences as cups and plates; but the existence and character of the electron depends on the interaction between humans and nature. That is to say that electrons, photons, etc. are artefactual.

This example from quantum mechanics is perhaps a little speculative. Let us therefore return to probability where we are on firmer ground in making the distinction between the fully objective and the artefactual. Let us begin with the standard example of tossing a biased coin. Here the probability of heads is objective, but clearly artefactual. The coin is a human artefact and its tossing is a human intervention carried out according to fixed rules. Exactly the same applies if we consider the probabilities in quantum mechanics which arise out of the repetition of some experiment. These probabilities are artefactual, and, if the analysis of the previous paragraph is correct, they have the same character as the micro-particles involved such as electrons. This, I think, sheds some light on Popper's wish to introduce the propensity interpretation of probability for use in quantum mechanics.

If now, however, we consider the probability of a radioactive atom disintegrating, we move away from the artefactual towards the fully objective, since the repetitions in this case (different atoms disintegrating) occur spontaneously in nature and do not require any human intervention. One difficulty, however, is that if we want to give a value to the probability of a radioactive atom of a particular type

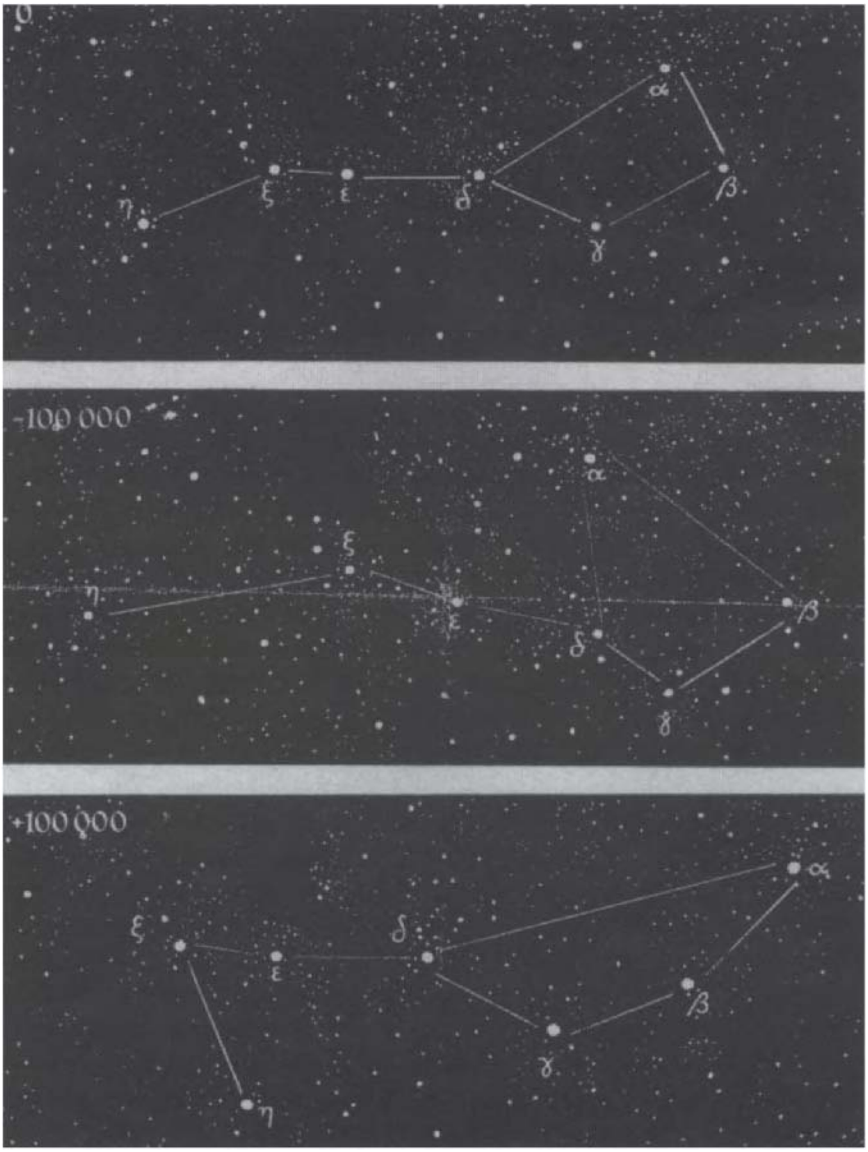


Figure 8.2 The Plough (or Big Dipper) today, 100,000 years ago and 100,000 years in the future

disintegrating, we have to specify a time period, say a year, and this time period might seem to be a human addition, making the probability more artefactual. There is a way of avoiding this problem. Radioactive disintegrations are Poisson processes (see Cramér, 1946:434–7, for some empirical evidence for this.) Now in a Poisson process the fundamental parameter λ say signifies that the probability of an event, e.g. a disintegration, in a small time interval δt is approximately $\lambda\delta t$. Thus λ can roughly speaking be considered as the probability of an event per unit time. The need for specifying a human relative time interval disappears, and λ can be considered as something fully objective.

We are now in a position to state explicitly our spectrum leading from the subjective to the objective, as applied to probability. It has four stages:

1 Subjective: Here probabilities represent the degrees of belief of particular individuals.
2 Intersubjective: Here probabilities represent the degree of belief of a social group which has reached a consensus.
3 Artefactual: Here the probabilities can be considered as existing in the material world and so as being objective, but they are the result of interaction between humans and nature. Many examples of artefactual probabilities can be given. Earlier we considered the case of Mr Smith aged 40 and whether he would live to be 41. We pointed out that Mr Smith could be classified in many different ways, and that each of these classificatory conditions yielded a different probability of his living to be 41. These probabilities are artefactual in a way similar to the constellations. Then again probabilities in coin tossing and other games of chance, as well as the probabilities associated with repeatable experiments in science, are artefactual.

   It is interesting to note that a criterion proposed by De Finetti classifies such artefactual probabilities as objective. De Finetti says:

> By denying any objective value to probability I mean to say that, however an individual evaluates the probability of a particular event, no experience can prove him right, or wrong; nor, in general, could any conceivable criterion give any objective sense to the distinction one would like to draw, here, between right and wrong.
>
> (1931a:174)

   Conversely one might say that if the evaluation of a probability can be shown to be right or wrong by experience, then that. probability can be regarded as objective. Now consider a typical artefactual probability such as the probability of getting heads with a particular biased coin tossed in a specified manner. If I judge this probability to be $^3/_4$, then, assuming as usual a falsifying rule, this evaluation can either be corroborated or refuted by a sequence of say two thousand tosses. Of course De Finetti would not have accepted methodological falsificationism or a falsifying rule, and this is how he would have defended his exclusively subjective position. If, however, we follow the majority of statisticians in adopting methodological falsificationism, then artefactual probabilities do become objective, according to De Finetti's own criterion of objectivity. De Finetti's criterion is also connected with the link suggested earlier (pp. 167–8) between objectivity and repeatable conditions. If a probability is associated with a set of repeatable conditions, we can test whether a conjectured evaluation of the probability is correct by repeating the conditions.

4 Fully objective: Finally we reach the highest grade of objectivity. Things should be considered fully objective which exist in the material world quite independently of human beings. As an example of the fully objective in the field of probability, we considered the probability per unit time of a particular radioactive atom disintegrating.

Earlier (pp. 18–20) we considered the Janus-faced character of probability, and distinguished the two kinds of probability as epistemological and objective. In the above spectrum 1 and 2, the degree of belief interpretations, are obviously epistemological, whereas 3 and 4 are objective. In Chapter 9 I will give some examples which are on the borderline between epistemological and objective. Since they fall between the cases 2 and 3 considered above, they will tend to make our spectrum yet more continuous.

Pluralist views of probability

Many of the authors we have considered so far claimed that their interpretation of probability applied to all uses of the concept. In other words, they advocated a monist view of probability. This was true of Keynes and De Finetti. It was also true in a sense of Von Mises. Von Mises did acknowledge that there was an ordinary language or common sense notion of probability which was not covered by his frequency theory. However, he claimed that this notion of probability was purely qualitative, and that the mathematical theory could not be applied to it. He thought that his frequency theory covered all the cases in which the mathematical theory of probability could validly be applied.

In contrast to these monist views of probability, there are pluralist views, according to which the mathematical calculus has a number of different interpretations each of which is valid in a particular area or context. In this section I will consider briefly this pluralist position. I will begin by discussing Ramsey's views on the question, since he seems to have been the first twentieth-century thinker to have advocated pluralism regarding probability.

Ramsey's position, known as the *two-concept view,* is stated in the Foreword to his 1926 paper as follows:

> In this essay the Theory of Probability is taken as a branch of logic, the logic of partial belief and inconclusive argument; but there is no intention of implying that this is the only or even the most important aspect of the subject. Probability is of fundamental importance not only in logic but also in statistical

> and physical science, and we cannot be sure beforehand that the most useful interpretation of it in logic will be appropriate in physics also. Indeed the general difference of opinion between statisticians who for the most part adopt the frequency theory of probability and logicians who mostly reject it renders it likely that the two schools are really discussing different things, and that the word 'probability' is used by logicians in one sense and by statisticians in another. The conclusions we shall come to as to the meaning of probability in logic must not, therefore, be taken as prejudging its meaning in physics.
>
> (Ramsey 1926:157)

Ramsey here suggests that the meaning of probability in logic, obviously taken to include inductive as well as deductive logic, may be different from its meaning in statistical and physical science. Such a position was definitely advocated by Carnap (1950:19–51). Carnap spoke of two concepts of probability, which he called probability$_1$ and probability$_2$. Probability$_1$ was probability as used in logic, and probability$_2$ was probability as used in statistical and physical science. Carnap advocated a frequency interpretation for probability$_2$, but for probability$_1$ he favoured, at least at first, a logical interpretation more similar to that of Keynes than to Ramsey's subjective approach. Indeed Carnap criticises what he calls 'psychologism in inductive logic' (1950:42–51).

Ramsey has some further things to say about the two-concept view in his 1926 paper. He argues that the mathematical calculus of probabilities can be given a frequency interpretation in terms of class ratios. Then, after introducing his alternative interpretation in terms of partial belief, he makes the following comment:

> ... we saw at the beginning of this essay that the calculus of probabilities could be interpreted in terms of class-ratios; we have now found that it can also be interpreted as a calculus of consistent partial belief. It is natural, therefore, that we should expect some intimate connection between these two interpretations, some explanation of the possibility of applying the same mathematical calculus to two such different sets of phenomena.
>
> (Ramsey 1926:187)

Ramsey here poses an important problem for anyone who advocates a pluralist view of probability. Such a person has to show how it is that the same mathematical calculus can have different interpretations, and how these different interpretations are related. Ramsey is dealing with two interpretations – a frequency interpretation and a degree of partial belief interpretation. He answers his own question as follows:

> Nor is an explanation difficult to find; there are many connections between partial beliefs and frequencies. For instance, experienced frequencies often lead to corresponding partial beliefs, and partial beliefs lead to the expectation of corresponding frequencies in accordance with Bernoulli's Theorem. But neither of these is exactly the connection we want; a partial belief cannot in

> general be connected uniquely with any actual frequency, for the connection is always made by taking the proposition in question as an instance of a propositional function. What propositional function we choose is to some extent arbitrary and the corresponding frequency will vary considerably with our choice. The pretensions of some exponents of the frequency theory that partial belief means full belief in a frequency proposition cannot be sustained. But we found that the very idea of partial belief involves reference to a hypothetical or ideal frequency; supposing goods to be additive, belief of degree *m*/*n* is the sort of belief which leads to the action which would be best if repeated *n* times in *m* of which the proposition is true; or we can say more briefly that it is the kind of belief most appropriate to a number of hypothetical occasions otherwise identical in a proportion *m*/*n* of which the proposition in question is true. It is this connection between partial belief and frequency which enables us to use the calculus of frequencies as a calculus of consistent partial belief. And in a sense we may say that the two interpretations are the objective and subjective aspects of the same inner meaning, just as formal logic can be interpreted objectively as a body of tautology and subjectively as the laws of consistent thought.
>
> (1926:187–8)

Ramsey's view as here expressed of the connection between frequency and partial belief is an attractive one. Particularly striking is his claim that 'the two interpretations are the objective and subjective aspects of the same inner meaning' (1926:188). Nonetheless, our earlier discussion (pp. 119–25) shows that Ramsey's account is too simple.

Let us return to our example of trying to assess the probability of Mr Smith, aged 40, living to be 41. One difficulty, which we called the 'reference class problem', is that Mr Smith can be classified under a number of different conditions. He can for example be considered as a man aged 40, as an Englishman aged 40, as an Englishman aged 40 who smokes two packets of cigarettes each day, and so on. Each set of conditions will give a different sequence of repetitions in which the frequency of those who survive to their 41st birthday will be different. To which frequency then do we relate our partial belief in the proposition that Mr Smith will live to be 41? As a matter of fact, Ramsey himself expresses just this difficulty in different terms as follows:

> '... a partial belief cannot in general be connected uniquely with any actual frequency, for the connection is always made by taking the proposition in question as an instance of a propositional function. What propositional function we choose is to some extent arbitrary and the corresponding frequency will vary with our choice.'
>
> (1926:188).

However, it seems to me that this problem vitiates the account of the matter which he gives a few lines later when he says: 'belief of degree *m*/*n* is ... the kind of belief most appropriate to a number of hypothetical occasions otherwise identical in a proportion m/n of which the proposition in question is true.' (1926:188). As we argued earlier (pp. 161–2), there cannot be occasions which are identical in any absolute sense, but only occasions which are identical in certain respects. But how do we choose these respects? This raises the reference class problem once again, since different respects will produce different frequencies.

Of course the reference class problem can to some extent be overcome by classifying an individual event as a member of the narrowest reference class for which statistical data are available (if there is such a class). Thus in our earlier example, we should certainly prefer to classify Mr Smith as a member of the class of those Englishmen aged 40 who smoke two packets of cigarettes each day, *provided* we have reliable frequency data for this class. But this device of the narrowest reference class is not sufficient to link subjective probabilities to frequencies as we saw from the Francesca argument.

In the case of Francesca the narrowest reference class was that of 16-year-old Romans who possessed a motor scooter, and the frequency in this class (f say) was of those who had a road accident. Now Francesca's argument was that being more sensible and competent than the average 16-year-old Roman, she would drive her scooter with more care and attention, so that the probability of her having an accident was less than f. Now for those who knew her character, Francesca's claim to be more sensible and competent than her peers did seem to be justified, and so her argument appeared to be correct. The conclusion here was that the degree of belief in her having an accident should be taken as different from the relevant frequency f. Indeed, as we saw, Keynes (1921:322) argued that it is very often unsatisfactory to base our probabilities for single events on some related statistical frequency, because in so doing we may well be neglecting some relevant non-statistical information about the specific case, which could lead us to making the probability either greater or less than the frequency. All this shows the inadequacy of Ramsey's account of frequency and degree of belief as: 'the objective and subjective aspects of the same inner meaning' (1926:188).

Ramsey's article of 1926 was mainly about the use of probability in logic as degree of partial belief. He intended, however, to write a further chapter about probability in statistical and physical science. Unfortunately, because of his early death on 19 January 1930, only fragments about this topic survive, and it is difficult to reconstruct what his view would have been. A scholarly and plausible account based on the surviving fragments has been given by Galavotti (1994, 1999). The 1999 paper is also interesting because of her attempt to develop a new approach to the question through a synthesis of the views of Ramsey and De Finetti. Here, however, I will not attempt any further discussion of Ramsey's position, but rather give a short account of my own views on the question.

Looking back over the chapters of this part of the book, it seems to me that the following conclusions can be drawn regarding possible interpretations of probability. First of all the classical interpretation, though satisfactory for games of chance, is not adequate for all the modern applications of probability. It must therefore be regarded as superseded. The logical interpretation still has contemporary advocates, but the difficulties connected with the Principle of Indifference seem to me to be fatal to the theory. The Principle of Indifference undoubtedly leads to paradoxes, and, although there are ingenious resolutions of some of these, there is no general method of eliminating them all. Anyone using the Principle of Indifference can never be sure if and when it is going to give rise to a contradiction. The only safe strategy is to abandon the principle altogether, and this means giving up the logical interpretation – at least in its traditional form. By contrast, the subjective interpretation and its off-shoot the intersubjective interpretation seem to me quite valid interpretations of the mathematical calculus of probability. The identification of degree of belief with a betting quotient and the Dutch book argument can and have been criticised, but they seem to me sufficiently realistic and convincing to give a sound foundation to subjective and intersubjective probabilities. On the other hand, De Finetti's attempt to reduce all probabilities to subjective probabilities via his 'exchangeability reduction' seems to me to fail (see the criticism pp. 77–83). Thus I would argue that an objective interpretation of probability is needed in addition to the subjective and intersubjective, and so I am committed to a pluralist view of probability.

Turning now to objective interpretations of probability, it seems to me impossible to deny that Von Mises' frequency theory is a valid interpretation of the probability calculus. The theory is provably consistent relative to classical mathematics, and its frequency interpretation of probability satisfies the Kolmogorov axioms with finite additivity. On the other hand, the propensity theory as developed in Chapter 7 seems definitely to be superior to Von Mises' theory on a series of points. The propensity theory is based on a non-operationalist view of conceptual innovation which explains conceptual innovation in the natural sciences better than Von Mises' operationalism; the propensity theory eliminates all the problems about infinite collectives, and, by introducing a falsifying rule for probability statements, gives an account of the relations between probability and frequency which agrees very well with standard statistical practice; the propensity theory eliminates Von Mises' introduction of the two separate concepts of randomness and independence by reducing both to independence; the propensity theory by associating probabilities with repeatable conditions rather than collectives allows for a wider range of applications of the calculus; the propensity theory fits in better with the Kolmogorov axioms and the modern mathematical approach to probability using measure theory, since it allows probability to be introduced as an undefined concept; and so on. Taking all these points together, I think we can definitely say that the propensity theory has superseded the frequency theory.

So the conclusion of all this discussion is that there are three currently viable interpretations of probability: the subjective, the intersubjective and the propensity. Interestingly, these correspond to what Fleck calls the three factors in cognition when he writes of: 'The three factors involved in cognition – the individual, the collective, and objective reality (that which is known)' (Fleck 1935:40). Now regarding these three interpretations we can repeat Ramsey's observation that 'It is natural ... that we should expect some intimate connection between these ... interpretations, some explanation of the possibility of applying the same mathematical calculus to ... such different phenomena.' (1926:187). Indeed for us, as for any proponent of a pluralist view of probability, it is necessary to explain the connection between the various interpretations.

The connection between subjective and intersubjective probability is straightforward since the latter is just the extension of the former from individuals to groups. The key problem is then to explain how the subjective interpretation of probability is connected to the objective propensity interpretation. My view is that the objective propensity interpretation should be taken as fundamental. Experimental investigations by gamblers using apparatus such as coins, dice, roulette wheels, etc. produced a mass of empirical material concerning random phenomena. In particular, such phenomena were found to obey two rough empirical laws – the Law of Stability of Statistical Frequencies and the Law of Excluded Gambling Systems. The mathematical theory of probability was developed to explain and render more precise these laws. Later it was extended to explain statistical phenomena in a wide variety of different areas from radioactivity to genetics. The basic interpretation of probability theory is thus as a mathematical science of randomness, and the theory's success in explaining (and rendering more precise) a mass of empirical material is what confirms its axioms, and justifies us in accepting them.

The subjectivists have shown that the mathematical calculus can be extended to deal with degree of belief in particular events. The connection between the two interpretations occurs in the area of games of chance. If a gambler is betting that the next roll of a die will give 5 say, what is important is his or her betting quotient on that particular roll. However, background knowledge will in general induce him or her to put this betting quotient equal to the objective propensity of the die yielding 5. Since betting quotients are equal to propensities in this particular case, it is not perhaps so surprising that they should obey the same mathematical calculus. However, this equality of betting quotients and propensities only really applies in the simple case of games of chance. If we are considering which horse will win a race, or even whether a particular person will have a road accident in the next five years, the betting quotient may well diverge from the propensity, or indeed it may be impossible to define any objective propensities. Thus the subjective interpretation of probability, while connected with the objective at one point, does genuinely extend the probability calculus to cases with which the objective interpretation cannot deal. This point of view actually agrees with what De Finetti says in the following passage:

> It would not be difficult to admit that the subjectivistic explication is the only one applicable in the case of practical predictions (sporting results, meteorological facts, political events, etc.) which are not ordinarily placed in the framework of the theory of probability, even in its broadest interpretation. On the other hand it will be more difficult to agree that this same explanation actually supplies rationale for the more scientific and profound value that is attributed to the notion of probability in certain classical domains, ...
>
> (1937:152)

Of course, De Finetti thinks that the subjective interpretation can be extended to cover these 'classical domains', and this is where I disagree with him (see the criticism of his exchangeability reduction pp. 77–83). I would argue that probability does really have a 'more scientific and profound value' in these domains, and this is what is most basic and important for the mathematical theory of probability. On the other hand, the subjectivists have shown how the use of the probability calculus can be extended from these classical domains to cases of 'practical prediction'. This extension is an important achievement.

The essential difference between the objective and subjective interpretations is in my view the following. In the objective interpretation, probabilities are associated with repeatable conditions which have independent outcomes. Since the conditions are repeatable, it is possible, if we adopt a falsifying rule, to test our probability ascriptions and either confirm or refute them. It is this characteristic which makes these probabilities objective. Subjective probabilities, on the other hand, are appropriate for singular events, either where no repeatable conditions can be easily defined, as in the case of horse races, or where such repeatable conditions as can be defined do not express all our knowledge relating to the individual event, as in the case of considering whether a particular person will have a road accident in the next five years. As I pointed out in earlier (pp. 167–8), it is possible to extend the Kolmogorov axioms to a formalism in which the assumption of repeatable conditions with independent outcomes is made explicit. This extension actually characterises the objective propensity interpretation axiomatically, and differentiates it from the subjective interpretation. However, the standard Kolmogorov axioms can be interpreted both subjectively and objectively.

This concludes my account of the various interpretations of probability and of how they are related. In the final chapter of the book, I will illustrate the pluralist view of probability by arguing that different interpretations of probability are appropriate for the natural sciences and for the social sciences.

9 An example of pluralism

Differences between the natural and social sciences

In the previous chapter I argued for the existence of several different, though interconnected, notions of probability which apply in different contexts. I suggested that these could be arranged in a series running from the subjective to the fully objective, but that it was still convenient to divide these according to the 'Janus-faced' character of probability into epistemological and objective interpretations. In this last chapter of the book, I want both to illustrate and to reinforce this pluralist view of probability by arguing that there are two broad areas of intellectual study which require different interpretations of probability. More specifically the thesis of the chapter will be that an epistemological notion of probability is appropriate for the social sciences, whereas an objective notion is appropriate for the natural sciences. Although this thesis is intended to apply to all the social sciences, I will concentrate on the question of interpreting probability in economics, since the rôle of probability in economics has been much discussed. It is worth noting in this connection that most of the principal advocates of the epistemological interpretation of probability (Keynes, Ramsey, De Finetti) were concerned with the application of probability in economics, and that most of the principal advocates of the objective interpretation of probability (Von Mises, Fisher, Neyman, Popper) were concerned with the application of probability in the natural sciences (physics and biology).

In the section 'General arguments for interpreting probabilities in economics as epistemological rather than objective', I will present some general arguments for interpreting probabilities in economics as epistemological rather than objective. As already observed, this implies a difference between the natural and social sciences. It turns out moreover that the arguments of this section have some features in common with the arguments which Soros presents (1987/94) for his thesis that the social sciences differ from the natural sciences. I will accordingly expound some of Soros's arguments in the section 'Soros on the difference between the natural and social sciences', and note their similarity with the arguments of the first section. Having in this way strengthened the thesis that there is an important difference between the social and natural sciences, I will use it to try to resolve an apparent contradiction which emerged earlier in the book (see the beginning of Chapter 7). In Chapter 4 on the subjective theory, I endorsed Ramsey and De Finetti's operationalist definition of degree of belief in terms of betting quotients as providing a satisfactory foundation for the subjective interpretation. Yet in Chapter 7 on the propensity theory, I criticise the use of operationalism in the natural sciences and develop a non-operationalist theory. It looks therefore as if I am both adopting and rejecting operationalism. In the light of the thesis of this chapter, however, the contradiction is easily resolved. In the last section I will argue that operationalism is appropriate for the social sciences but not for the natural sciences.

General arguments for interpreting probabilities in economics as epistemological rather than objective

A useful way into this discussion is through the consideration of Lad's interesting article of 1983. In this paper, Lad argues strongly against the objective interpretation of probability, which he rejects *in toto.* I will adopt a position which partially agrees and partially disagrees with Lad. My claim will be that Lad's arguments do rule out an objective interpretation of probability in economics, but that they are not a valid criticism of an objective interpretation of probability in the natural sciences. Lad challenges the objective interpretation of probability presented by Gnedenko (1950). However, this interpretation is in fact quite similar to the propensity interpretation given in Chapter 7. Both regard objective probabilities as being associated with repeatable conditions whose outcomes are independent. Earlier (p. 184), I argued that the propensity theory is the best objective interpretation of probability currently available. Throughout this section therefore I will take the objective interpretation to be the propensity theory. This will simplify the exposition, but it is not necessary for the arguments. Those who still prefer Von Mises' frequency theory need only replace, in what follows, consideration of independent trials of a repeated condition by consideration of random collectives. The arguments will go through just the same.

Lad begins his criticism by stating an important assumption of Gnedenko's objectivism as follows:

> ... an event A is said to have a probability relative to condition C if: a) It is possible, at least in principle, to set up an unlimited number of mutually independent trials of A under the same repeated condition C; ...
>
> (1983:290)

This of course is the key feature of the propensity theory of Chapter 7, according to which objective probabilities are associated with a set of repeatable conditions whose repetition yields independent trials. Lad denies that such a set of conditions can be found:

> The assertion of the repeatability of conditions of an experiment is fundamentally a manifestation of a metaphysical mode of thinking ... The two experiments are clearly completely different events, distinct in at least time or space, and an infinity of other circumstances as well.

> Surely the dialectic process of evolving nature does not admit the possibility of repetitions of identical circumstances, necessary in principle for the probability of Gnedenko's construction.
>
> (1983:291)

Lad is of course correct in asserting that two experiments are different events which will differ in several circumstances. On the other hand, it may be possible to find a set of conditions C such that (1) although the repetitions of C do differ in several circumstances, we can neglect these variations for the purpose we have in hand; and (2) the repetitions have so little influence on each other that they can be regarded as effectively independent. In other words, we may be able to produce repetitions of a set of conditions C, which, from a practical point of view, can be regarded as independent. Surely this is the case whenever the experimental method is applied successfully, as it so often is in the natural sciences. As far as economics is concerned, I agree with Lad that it may not be possible to specify in a satisfactory manner a sequence of independent repetitions.

Hicks (1979:103–22) has expressed a point of view similar to the one defended in this section. Hicks begins by distinguishing two interpretations of probability:

> It is the frequency theory which has become orthodox; most modern works on statistical mathematics take it as their starting point. The chief proponents of the alternative approach have been Keynes, in his *Treatise on Probability* (1921) and Harold Jeffreys, in his *Theory of Probability* (1939).... It is ... significant that Keynes, the modern economist who has thought most deeply on these matters, was a proponent of the alternative theory. I have myself come to the view that the frequency theory, though it is thoroughly at home in many of the natural sciences, is not wide enough for economics.
>
> (1979:105)

Hicks is here contrasting two interpretations of probability – the frequency and the logical. Our own framework is wider since we distinguish *objective* theories of probability (which include both the *frequency* and the *propensity* interpretations) from *epistemological* theories (which include the *logical, subjective* and *intersubjective* interpretations). However, when Hicks says that the frequency theory is appropriate for the natural sciences and the logical theory for economics, he is coming close, within his own framework, to the view of this section that objective probabilities are appropriate for the natural sciences, and epistemological probabilities for economics. This is made clearer by the following passage from Hicks:

> According to the frequency theory, probability is a property of *random experiments*....
>
> There clearly are cases, important in economics, in which we speak of probability in another sense. Cramér ... writing the chapter of his book in which it occurs at the end of 1944 ... gives, as an example ... the probability

> that the European war would come to an end within a year. This was a probability which, at that date, most people would have assessed to be a high one. But it is quite clear that it does not fall within the frequency definition; it is not a matter of trials that could be *repeated.*
>
> We cannot avoid this other kind of probability in economics. Investments are made, securities are bought and sold, on a judgment of probabilities. This is true of economic behaviour; it is also true of economic theory. The probabilities of 'states of the world' that are formally used by economists, as for instance in portfolio theory, cannot be interpreted in terms of random experiments. Probability, in economics, must mean something wider.
>
> (1979:105–7)

To explore this question further it will be useful to compare a typical situation in economics with one in physics. For economics, let us consider the attempt to build a model of a capitalist economy. For physics, we shall take the kinetic theory applied to the gas in a container. These examples have been chosen because there is a certain structural similarity between them. The economy consists of a set of agents performing all sorts of actions, whereas the gas consists of a set of molecules moving around with different velocities. Despite these similarities, I will argue that the two cases differ in important respects.[1]

The key difference seems to be this. The molecules have no knowledge, consciousness or volition, and, apart from the occasional collision, move to a first approximation independently of each other.[2] The economic agents, however, do possess knowledge, consciousness, desire and will. Moreover, their actions, far from being independent, are characterised by reaction to each other's actual or expected decisions.

So, on the one hand, we can introduce a degree of belief interpretation in the economic case, although molecules obviously do not have beliefs, whereas, on the other hand, the independence assumption which is basic to the objective interpretation of probability does not seem to apply in economics. This all points to the conclusion that we need an objective interpretation of probability for the kinetic theory of gases and an epistemological interpretation of probability for the analysis of a capitalist economy.

Against this, it might be argued that we have exaggerated the difference between the two cases by claiming that, while the molecules are quite independent, the economic agents strongly interact. In fact, it might be said, more advanced treatments of the kinetic theory of gases allow for interactions between the molecules and there might be a parallel here to the interactions between economic agents.[3] In order to answer this point, we shall have to examine in a little more detail the rôle of probability in the kinetic theory of gases.

One of the basic results of the kinetic theory of gases is Maxwell's law of distribution of velocity, first obtained by Maxwell in 1860. Let us consider a volume V of gas at temperature T and suppose that the gas contains n molecules. The problem is to calculate the number n_v of molecules which have a velocity between v and $v + dv$. Maxwell, starting with some quite plausible probabilistic assumptions, obtained the law:

$$n_v = \lambda v^2 \exp(-\mu v^2) \text{ where } \lambda, \mu \text{ are constants} \tag{9.1}$$

This law was tested out experimentally by Stern in 1920. I will sketch his method, full details of which are to be found in Fraser (1931:60–74).

A sample of gas of volume V and temperature T is prepared in a chamber A (Figure 9.1). Some of the molecules are allowed to escape though an opening O_1 into a second chamber B. A second opening O_2 narrows down these molecules into a molecular ray in a third chamber C. If the molecules in A follow Maxwell's law of distribution of velocity, those in the molecular ray will follow the related law:

$$n_v = \lambda v^3 \exp(-\mu v^2) \tag{9.2}$$

Stern used some ingenious methods to measure the distribution of velocities in the molecular ray, and in fact obtained good agreement with the predictions of Maxwell's law, i.e. with Equation 9.2.

The point to note here is that a sequence of independent repetitions of Stern's experiment is perfectly possible. The experiment can be performed in the same laboratory on different days, or in different laboratories on the same day, and these repetitions will be independent. Thus the conditions noted earlier as necessary for the introduction of objective probabilities are satisfied; and indeed we can take the probabilities in the kinetic theory of gases to be objective.

The next point to observe is that this continues to hold even if we complicate the kinetic theory of gases by introducing interactions between the molecules. We have argued that independence is necessary for objective probabilities, but this independence need not be the independence of the various movements of the molecules, for, even if there is interaction between the molecules, we can still have independent samples of the same gas all having the same volume *V* and temperature *T.* These independent samples can be prepared at the same place at different times, or at different places at the same time. Is there anything analogous to these independent samples of gas in the case of capitalist economies? I will next argue that there is not, and that this prevents the introduction of objective probabilities in the economics case.

Controlled experiments are, of course, extremely difficult in economics. Can we, however, use the observations of behaviour and performance of economic

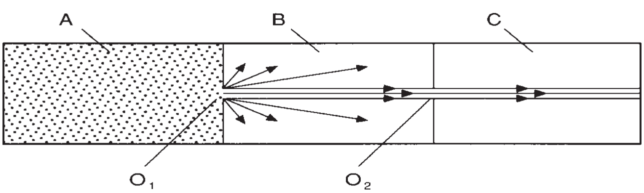


Figure 9.1 Stern's apparatus for testing Maxwell's law of distribution of velocities experimentally

systems as samples of independent repetitions of conditions similar to the ones related to gases in Stern's experiment? The different samples could be taken from either (1) data related to the same economic system at different times, or (2) data related to different economic systems at a similar stage of development (e.g. France and Germany).

In the first case, if the samples refer to 'snapshots' of the economy which are too close together in time, it is hard to maintain that the more recent performance is not influenced by that of the previous periods; thus the independence of the samples cannot be maintained. If the samples relate to historical periods far enough from each other to render the assumption of independence plausible, one is unlikely to get homogeneous samples; thus invalidating the 'experiment'. In the second case the use of a sample of cross-section data would still not give independence as economic systems tend to be integrated in terms of trade and production, and particularly as the flow of information from one country is likely to affect the behaviour of agents in others.

It seems therefore impossible to introduce a satisfactory notion of an independent repetition of the state of an economy, and we cannot therefore use objective probabilities in economics. It might be objected that we can overcome the difficulties caused by lack of independent repetitions by the method of random sampling. Hicks considers a case of this sort in the following passage:[4]

> When we are looking for a conclusion that is to be derived from sampling (as for instance in the study of family budgets) it is possible to take steps to ensure that the sample is random, or at least fairly random; we then have a right to make use of sampling theory, which (as explained) is a branch of the probability calculus.
>
> (1979:120)

The problem here is that while we do have randomness (or independent repetitions) in such cases, they are *introduced by the sampling procedure* and do not occur in the reality actually under study. Correspondingly it is inappropriate to introduce objective probabilities in such cases since all we have are fixed and definite (though possibly unknown) frequencies.

This can be seen by considering the following simple example. Suppose we have six hemispherical holes cuts in a straight line in a piece of wood and numbered 1–6. Suppose four white balls are placed in holes 1–4, and two black balls in holes 5 and 6. No one would say that this fixed, and perfectly definite arrangement, contains any chance element or objective probabilities.

We can however introduce random sampling in the following way: roll a die, and, if the result is *n,* note the colour of the ball in hole *n.* This would produce a random sequence of 'white' and 'black'. Someone who could not see the balls themselves, but had access to the results of the random sampling device, would be able to infer the proportion of white to black balls in the hidden arrangement. This would not show that the hidden arrangement contained objective probabilities. The objective probabilities in the example are all introduced by the random sampling procedure (rolling the die).

Hicks's example of the family budgets (in England say) is exactly the same except that the numbers involved are larger. At any given time, the proportion of families in England with budgets in a given range is a perfectly definite, though perhaps unknown, number. It is a frequency, not an objective probability. By taking a random sample we can use probability calculations to estimate this unknown number, but the objective probabilities in these calculations are introduced by the random sampling procedure and do not occur in the reality under study. So Hicks's example (and other similar examples) does not show that objective probabilities can be validly introduced in economics.

That ends my arguments for interpreting probabilities in economics as epistemological rather than objective. By extension we can conclude that the epistemological interpretation of probability is appropriate for the social sciences in general. This therefore is a point on which the social sciences differ from the natural sciences, since probabilities in the natural sciences are objective.

Problems are raised for this view which distinguishes between the natural and social sciences by cases which lie on the borderline of these two disciplines. An example is the growth or decline of human populations. Now biology studies the growth or decline of animal populations, and the problem in the human case may seem fairly similar. After all, humans reproduce sexually just like other mammals, and there is certainly a biological basis to human reproductive behaviour. In this respect the human case is similar to the animal case, and one might argue that the problem of the growth or decline of human populations is essentially a biological problem, and so part of the natural sciences. Against this, however, it cannot be denied that social factors of a political and/or economic character do play a part in determining the growth or decline of human populations. An obvious recent example is the one-child policy which is in force at present in China. This was enacted politically with the object of enabling the standard of living of the average Chinese person to rise, and it is enforced by the State. There is nothing biological about all this, but the policy is undoubtedly having an effect on the size of the Chinese population. Sometimes economic factors operate in a less obvious way to have much the same effect. Thus it seems to be a general law that as countries reach higher levels of industrialisation, the birth rate drops, even though the average family has become richer and could presumably afford to have more children than before. Thus the birth rate in Italy at the moment is actually lower than in China, even though there is no one-child family policy, and even though Italy in the past was noted as a country whose people loved children and large families.

Medicine is another example of a subject which lies on the interface between the social and natural sciences. At one level the human body can be considered as a complicated biochemical mechanism; and diseases can be considered as caused by the malfunctioning of this mechanism owing either to some internal problem or to the invasion of external entities such as pathogenic bacteria. From this point of view medicine is just a branch of the natural sciences. On the other hand, there are well-authenticated instances of the influence of the mind over the body. For example, placebos do have a curative effect even though they are chemically neutral. Here then psychology, one of the social sciences, enters the picture. Sociology is important as well as psychology. For example, the improvement of hygiene in society through the construction of drains and educating people in habits of cleanliness brings about a dramatic reduction in disease. Or again an economic slump with consequent mass unemployment brings in its train an increase in all kinds of diseases.

Now if epistemological probabilities are appropriate for the social sciences, and objective probabilities for the natural sciences, what are we to say about borderline subjects such as medicine or population studies? My answer is that this reinforces the view for which I have argued previously (pp. 175–80) that there is something in the nature of a continuous spectrum of interpretations of probability running from the completely subjective to the fully objective. The distinction between epistemological and objective interpretations of probability is still a useful one, but it must be remembered that it is the drawing of a line in what is to a large extent a continuum. Thus the existence of borderline cases should not surprise us.

Let us consider our two borderline cases a little more closely. It seems to me that human population studies lie more in the social than the natural sciences. Historical data show that rates of human population growth or decline vary enormously with changes in the social and economic situation, while presumably the biological character of human beings remains relatively fixed. Thus here the best approach to the problem is to focus principally on social, political and economic causes, and to regard the biological basis as secondary. Things are the other way round in the case of doctors treating patients. The probability of a patient's having a particular disease will depend on the social, political and economic situation of the country or region, but this probability can largely be relegated to background knowledge and will not have a great effect on the doctor's treatment of a specific case. As for psychological effects such as the placebo effect, or the effect of patients' beliefs, morale, stress etc. on the probability of their catching a disease and the difficulties of recovery, these have been studied statistically and their strength and limits are quite well known. So the doctor can focus principally on the patient's bodily condition as the key element in the disease and can regard the social and psychological factors as secondary. Thus, I would argue that, as far as the question of treating patients is concerned, medicine belongs primarily to the natural sciences, so that objective probabilities should have an important rôle.[5]

So far I have examined the differences between the natural and social sciences primarily from the point of view of probability theory. I now want to look at the question in more general terms, and, more specifically, to examine Soros's views on this question. These views are based on his analysis of financial markets, a social phenomenon which is just about as far from the natural sciences as it is possible to get. In the non-human world of nature, there is nothing like that human social creation, money; and still less is there anything like a market for stocks or derivatives. It follows that financial markets are a pure case of the social, and the analysis of them should enable us to pick out some of the key respects in which the social sciences differ from the natural sciences. This will be the theme of the next section.

Soros on the difference between the natural and social sciences

Soros was a student of Popper's at the London School of Economics and remains a great admirer of much of Popper's philosophy. However, there was one point on which he disagreed with Popper, and which he explains as follows:[6]

> I was greatly influenced ... by Karl Popper's ideas on scientific method. I accepted most of his views, with one major exception. He argued in favor of what he called "unity of method" – that is, the methods and criteria that apply to the study of natural phenomena also apply to the study of social events. I felt that there was a fundamental difference between the two: the events studied by the social sciences have thinking participants; natural phenomena do not. The participants' thinking creates problems that have no counterpart in natural sciences.
>
> (Soros 1987:11–12)

Consider for example a group of natural scientists studying astronomy. The stars, planets and comets do not think, nor could they be influenced by any theories which the group proposes. The situation is quite different in the case of a group of social scientists studying financial markets. The participants in these markets do think and are trying to understand the markets in which they are participating. Even if the group of social scientists is not itself participating in the financial markets, the theories which it proposes might well influence future changes in these markets. The two cases are then very different. As Soros puts it:

> Natural scientists have one great advantage over participants; they deal with phenomena that occur independently of what anybody says or thinks about them. The phenomena belong to one universe, the scientists' statements to another. The phenomena then serve as an independent, objective criterion by which the truth or validity of scientific statements can be judged. Statements that correspond to the facts are true; those that do not are false. To the extent that the correspondence can be established, the scientist's understanding qualifies as knowledge. We do not need to go into the various difficulties that stand in the way of establishing this correspondence. The important point is that scientists have an objective criterion at their disposal.
>
> By contrast, the situation to which the participants' thinking relates is not independently given: it is contingent on their own decisions. As an objective criterion for establishing the truth or validity of the participants' views, it is deficient. It does provide a criterion of sorts: some expectations are validated by subsequent events, others are not. But the process of validation leaves something to be desired: one can never be sure whether it is the expectation that corresponds to the subsequent event or the subsequent event that conforms to the expectation. The segregation between thoughts and events that prevails in natural science is simply missing.
>
> (1987:32–3)

One interesting idea here is that expectations can sometimes predict the outcome correctly because they influence the outcome. Thus investors might expect for no very good reason that a share's price will increase. They therefore buy the share in large numbers causing its price to increase. Soros elaborates on this point in the course of criticising the efficient market theory, which naturally he does not accept:

> The generally accepted view is that markets are always right – that is, market prices tend to discount future developments accurately even when it is unclear what those developments are. I start with the opposite point of view. I believe that market prices are always wrong in the sense that they present a biased view of the future. But distortion works in both directions; not only do market participants operate with a bias, but their bias can also influence the course of events. This may create the impression that markets anticipate future developments accurately, but in fact it is not present expectations that correspond to future events but future events that are shaped by present expectations.
>
> (Soros 1987:14)

Soros's demarcation between the natural and social sciences is in line with our discussion of the interpretation of probability in the two cases. In the previous section, we contrasted the molecules of a gas with the people in a capitalist economy. Molecules do not have thoughts and beliefs, whereas people do, and this was one reason why a degree of belief interpretation of probability was appropriate in the latter case, but not in the former. Our other main point was that in the case of the gas, independent repetitions were possible, but not in the case of the economy. This point is not so much stressed by Soros, but it seems to be implicit in what he says. Since a social system is composed of thinking participants, an independent repetition of any situation becomes difficult; for suppose a later situation is similar in some respects to an earlier one, the participants the second time round will know what happened on the previous occasion, and this is likely to affect the outcome of the later situation.

Soros regards Heisenberg's uncertainty principle in quantum mechanics as the closest analogy within the natural sciences of the features which he regards as characteristic of the social sciences. Even here, however, he thinks, quite rightly in my view, that the difference is considerable. In the case of Heisenberg's principle, it is only a question of *observations* influencing the subject matter. In the social sciences, it is *thoughts and beliefs as well as observations* which influence the subject matter. As Soros says:

> ... in quantum physics it is only the act of observation which interferes with the subject matter, not the theory of uncertainty, whereas in the case of thinking participants their own thoughts form part of the subject matter to which they relate. The positive accomplishments of natural science are confined to the area where thinking and events are effectively segregated. When events have thinking participants that area shrinks to the vanishing point.
>
> (1987:33)

The key concept which Soros uses to analyse social systems with thinking participants is that of *reflexivity.* The idea here is that the thinking participants in a social system study and analyse their social situation and form beliefs and theories about it. These beliefs and theories may be, indeed almost always will be, full of errors and misconceptions. Despite their inevitable defects these beliefs and theories influence the participants' actions, and so help to mould the way the social system develops. This development of the social system in turn influences the participants' beliefs and theories about it, and so on. The whole system evolves through a continuous process of interaction which Soros calls reflexivity. This is how he describes it:

> This process is fundamentally different from the processes that are studied by natural science. There, one set of facts follows another without any interference from thoughts or perceptions (although in quantum physics, observation introduces uncertainty). When a situation has thinking participants, the sequence of events does not lead directly from one set of facts to the next; rather, it connects facts to perceptions and perceptions to facts in a shoelace pattern. Thus, the concept of reflexivity yields a "shoelace" theory of history.
>
> It must be recognised that the shoelace theory is a kind of dialectic. It can be interpreted as a synthesis of Hegel's dialectic of ideas and Marx's dialectical materialism. Instead of either thoughts or material conditions evolving in a dialectic fashion on their own, it is the interplay between the two that produces a dialectic process.
>
> (Soros 1987:42–3)

Soros criticises neo-classical economics for failing to recognise reflexivity, and hence for producing a theoretical construction with little relevance to the real world. Whereas according to neo-classical economics, free markets have a built in tendency to move towards equilibrium, Soros denies that there is any such tendency, and even goes so far as to argue that markets tend towards excess and disequilibrium. This is how he puts it:

> Economic theory tries to sidestep the issue by introducing the assumption of rational behavior. People are assumed to act by choosing the best of the available alternatives, but somehow the distinction between perceived alternatives and facts is assumed away. The result is a theoretical construction of great elegance that resembles natural science but does not resemble reality. It relates to an ideal world in which participants act on the basis of perfect knowledge and it produces a theoretical equilibrium in which the allocation of resources is at an optimum. It has little relevance to the real world in which people act on the basis of imperfect understanding and equilibrium is beyond reach.
>
> (1987:12)

and again:

> It is almost redundant to criticize the concept of equilibrium any further. In the first chapter, I asserted that the concept is a hypothetical one whose relevance to the real world is open to question. In subsequent chapters I examined various financial markets as well as macro-economic developments and showed that they exhibit no tendency towards equilibrium. Indeed, it makes more sense to claim that markets tend towards excesses, which sooner or later become unsustainable, so that they are eventually corrected.
>
> (1987:317)

In 1994 Soros modified these views somewhat by arguing that, although reflexivity could arise at any time, in most situations it was sufficiently small to be ignored. In such situations neo-classical economics with its tendency to equilibrium could be applied. Sometimes however reflexivity becomes dominant, and then neo-classical economics becomes quite inappropriate. This is what he says:

> In *The Alchemy of Finance,* I put forward the theory of reflexivity as if it were relevant at all times. That is true in the sense that the two-way feedback mechanism that is the hallmark of reflexivity can come into play at any time, but it is not true in the sense that it is at play at all times. In fact, in most situations it is so feeble that it can be safely ignored. We may distinguish between near-equilibrium conditions where certain corrective mechanisms prevent perceptions and reality from drifting too far apart, and far-from-equilibrium conditions where a reflexive double-feedback mechanism is at work and there is no tendency for perceptions and reality to come close together without a significant change in the prevailing conditions, a change of regime. In the first case, classical economic theory applies and the divergence between perceptions and reality can be ignored as mere noise. In the second case, the theory of equilibrium becomes irrelevant and we are confronted with a one-directional historical process where changes in both perceptions and reality are irreversible. It is important to distinguish between these two different states of affairs because what is normal in one is abnormal in the other.
>
> (Soros 1994:6)

Although reflexive situations may occur now and then rather than all the time they are the situations which interest Soros, because they afford him the possibility of money making.

It might be thought that Soros developed his idea of reflexivity from his experience operating in the stock market and other financial markets, but this was not the case. As he says: '... I did not develop my ideas on reflexivity in connection with my activities in the stock market. The theory of reflexivity started out as abstract philosophical speculation and only gradually did I discover its relevance to the behavior of stock prices.' (Soros 1987:46) This brings us to the intriguing question of the influence of Popper's philosophy on Soros's business career. Soros himself has this to say on the subject:

> ... I want to acknowledge my indebtedness to Karl Popper's philosophy.... Karl Popper's philosophy has had a formative influence on my entire outlook on life. It has affected not only my thinking but also my actions. Strange as it may seem, it has made a tangible contribution to my business success ...
>
> (1992:1)

This is interesting because it runs counter to the popular idea of the unworldly and contemplative life of philosophy which is in sharp contrast to the practical nature of business. Could philosophy and business be more closely connected than is generally imagined? It is worth considering Soros's case to see whether this might be the case.

It seems that the key to Soros's business success was his theory of reflexivity. Soros (1987: Chapter 2) applies this theory to the stock market to produce a model which served him well in practice. Regarding this model, he says:

> The rudimentary model I have outlined above has proved extremely rewarding in my career as an investor. That may seem surprising because the model is so simple and it fits a well-trodden stock market pattern so well that one would expect every investor to be familiar with it. Yet, that is not the case. Why? Part of the answer must be that market participants have been misguided by a different theoretical construction, one derived from classical economics and, even more important, from the natural sciences....
>
> The first time I used the model systematically was in the conglomerate boom of the late 1960s. It enabled me to make money both on the way up and on the way down.
>
> (Soros 1987:55)

We see that Soros here points to a double advantage which he had over the average investor. On the one hand, he had developed a fairly realistic model based on reflexivity, whereas on the other hand the average investor having been trained in neo-classical economics was misguided by what Soros describes quite accurately as: '. . a theoretical construction ... that ... does not resemble reality.' (1987:12). Now Soros himself had been trained in neo-classical economics at the London School of Economics. Yet he had the independence of mind to reject that theory and develop more realistic models of markets. Perhaps it was here that Popper's philosophy helped, because Popper always stressed the need to criticise our theories and try to replace them with new theories which are better representations of reality.

That concludes my account of Soros's views. His arguments for an important difference between the social and natural sciences seem to me convincing and to reinforce the arguments given earlier (pp. 188–94). I will now use this thesis to resolve an apparent contradiction about operationalism which emerged earlier in the book.

Operationalism is appropriate for the social sciences, but not for the natural sciences

I have argued in this chapter and the previous chapter in favour of a pluralist view of probability which involves accepting both the subjective (and intersubjective) interpretation of probability, and also an objective propensity interpretation. We must now face up to a problem to which this pluralism gives rise, namely that the foundations of the subjective (and intersubjective) interpretations, on the one hand, and of the propensity interpretation, on the other, are radically different.

Let us begin with the subjective interpretation. This proceeds by identifying the degree of belief of a particular individual (Mr A say) with the rate at which Mr A would bet under specified conditions (Mr A's betting quotient). This is in effect an operational definition of degree of belief, and so we could say that the subjective theory (and its off-shoot the intersubjective theory) are based on operationalism. Indeed (as was pointed out p. 58), one of the best recent accounts of statistics using subjective probability, Frank Lad's 1996 book is entitled *Operational Subjective Statistical Methods,* and he explicitly appeals to the philosophy of operationalism in developing the foundations of subjective probability.

On the other hand, in developing a long-run version of the propensity theory of probability in Chapter 7, I explicitly criticised operationalism and advocated a non-operationalist theory of conceptual innovation. If then we are to accept, as our pluralism requires, both the subjective theory and the propensity theory of Chapter 7, we have apparently both to advocate and to repudiate operationalism. This is an awkward situation to say the least, and something needs to be done to resolve the contradiction.

One approach to the problem is to be found in Galavotti's 1995 paper 'Operationism, Probability and Quantum Mechanics'. Here Galavotti shows that the independent, but more or less contemporary, developments of subjective probability and of quantum mechanics both relied heavily on operationalist ideas. On the other hand, the pioneers of quantum mechanics such as Heisenberg and Born did not adopt the extreme subjectivism of De Finetti. This suggests that one might adopt a universal operationalist philosophy, but have an objective operationalist interpretation of probability in physics and a subjective operationalist interpretation of probability in other areas. The difficulty with this approach as far as I am concerned is that, if it is accepted, the objective operationalist interpretation of probability would be some form of the frequency theory rather than a propensity theory of the kind advocated in Chapter 7. Thus I prefer a different way of resolving the problem.

So far in this chapter I have presented arguments to the effect that there are fundamental differences between the social and natural sciences. Against this background, I want now to claim that operationalism is implicated in these differences, and that, in effect, operationalism is appropriate for the social sciences but not for the natural sciences. It is easy to see that this claim, if true, resolves our problem. The subjective theory of probability is concerned with measuring degrees of belief, and so belongs to psychology, one of the social (or human) sciences. So if operationalism is appropriate for the social sciences, it is appropriate for subjective probability. The propensity theory is concerned with interpreting probability in the natural sciences (e.g. physics and biology), and so, if operationalism is not appropriate for the natural sciences, it is not appropriate for the propensity theory. But are there any good reasons for supposing that operationalism is appropriate for the social sciences, but not for the natural sciences? This is the question to which we must now turn.

The whole problem will I believe be illuminated by introducing a new example which throws into sharp relief the issues involved here. This is the example of marking examination papers and classifying degrees. I will take as a specific case a degree with which I have been involved, namely the philosophy and mathematics undergraduate degree at King's College London. The students taking this degree do a mixture of philosophy and mathematics courses, which are all assessed by examination. In the philosophy exams, the students are asked to write three essays in 3 hours on three topics chosen from a list of about ten which cover the material of the course. The essays are then marked out of 100, and the total divided by three to give a mark out of 100 for the script as a whole. There are four grades: 70+ is a first, 60–69 an upper second, 50–59 a lower second and 40–49 a third. Below 40 is a fail. Now the thing to note here is that giving a philosophy essay an exact mark out of 100 is a somewhat arbitrary procedure. Of course everyone might agree that some essays are brilliant, some sound but uninspiring, some pretty mediocre and others positively bad. However, to go from this to saying that one essay is worth 47 and another 63 is a rather big step. Nonetheless, attempts have been made to introduce criteria so that the marking becomes less arbitrary. Each script is marked independently by two internal examiners, and, if these two examiners cannot agree through discussion, the issue is resolved by an external examiner. Although differences between the two internal examiners do indeed occur, it is perhaps more surprising that there is very often quite close agreement.

The undergraduate degree takes three years, and when it is completed the student will have taken a large number of exams for each of which he or she will have been awarded a mark. We now come to the next step which is that of giving the student a classification for the degree as a whole. This again will be first, upper second, lower second, third or fail. To produce an overall classification, it is obviously necessary to combine all the examination marks using some formula. The simplest idea would be just to take an average of all the student's marks. However, rightly or wrongly, this simple formula is not adopted. There are two arguments against it. First of all, it is thought that the examinations in the third year should count more than those in the second year, and those in the second year more than those in the first year. Thus a weighting is introduced. Second, it is thought to be unfair to a student that he or she should be brought down by a bad performance in one or two examinations, since these bad performances might have been due to an off day, or to an aversion to a particular subject or teacher. Thus, broadly speaking, the overall assessment is based on the best three-quarters of the student's marks.

Once again, however, matters are not so simple. Suppose the rule was adopted that only the best three-quarters of a student's marks were considered for the final classification. This might well lead some students to concentrate all their efforts on doing well on three-quarters of their courses and not bothering with the others on the grounds that they would not count anyway. To avoid such a response on the part of the students, some weighting is given to the worst one-quarter of their examination marks, though this is less than the weighting given to the best three-quarters.

It will be clear by now that the formula for combining a student's marks to give an overall degree classification must be quite complicated, and this is indeed the case. Moreover, King's College London has recently changed the formula for the undergraduate degree in philosophy and mathematics. Hitherto a formula known as the A-score has been used. However, after a great deal of discussion on committees during the academic year 1997–8 (similar to what has been given in the last few paragraphs, but more complicated), the college has introduced a new formula known as the I-score. This will gradually be phased in and eventually completely supersede the A-score. Now the interesting point to note is that a particular student, John Smith say, might be awarded a first for his degree as a whole on the basis of the A-score, but only an upper second on the basis of the I-score. This concludes my brief account of some methods used to mark examination papers and produce degree classifications. I now turn to the philosophical significance of these social procedures.

What I have been describing are operational procedures which have been laid down, and which enable marks and grades to be assigned both to individual examination papers and to a student's degree as a whole. It would, however, be rather implausible to claim that an examination paper or a degree had a real value *before* these procedures were introduced and that the procedures are only attempts to measure this pre-existing value. It hardly seems to make much sense to debate whether the A-score or the I-score best captures the real value of a student's performance, because it could certainly be questioned whether there was such a real value expressible by a numerical mark. In effect the numerical value of the degree is created by a convention which is chosen to be operationally applicable. This is not to say of course that the convention is wholly arbitrary. There is a background of rough qualitative agreement, and the method of numerical assessment has to be chosen so that it agrees with this background. However, this requirement leaves quite a lot of room for different A- and I-scores, etc.

Although there is undoubtedly considerable arbitrariness in the introduction of numerical marks and the division into classes, these procedures do really make a difference. One could say that they alter social reality, that human beings do, so to speak, get branded with a number which affects their position in society and life chances. Let us return to John Smith, who, we shall suppose, graduates in the last year in which King's College London's philosophy and mathematics degree is assessed by the A-score. He gets a first, and wants to go on to do a PhD. Let us suppose further that a first class degree is needed to be accepted for a PhD programme. John Smith is accepted and goes on to be acclaimed as the new Frank Ramsey. One year later Ann Jones completes her philosophy and mathematics degree. By chance she gets exactly the same marks on all the papers as John Smith, but now the overall degree is calculated by the I-score and she ends up with an upper second instead of a first. She too would like to do a PhD, but, because she failed to get a first, she is not accepted for the programme. So instead she founds an internet company and makes a large fortune. We see that the apparently arbitrary decision to substitute the I-score for the A-score could have the most profound effects on people's lives.

But why are numerical values and precise classes introduced for degrees at all? The answer is simple. There is strong pressure from employers for this to be done. Employers have to decide which graduates to hire, and a simple overall summary of a graduate's degree performance is helpful to them in making their decisions. Of course critics of modern society would object that the whole procedure is highly alienating. A complex multifaceted human being with all sorts of different abilities and weaknesses is reduced to a single number or grade in order to be slotted into a narrow and one-dimensional system. 'Surely', our critic might say, 'this is alienation.' I would be inclined to agree, while adding the pessimistic note that such alienation may be necessary in the present state of society.

Returning now to probability, we can see that assigning values to an individual's degrees of belief by betting quotients is a very similar procedure to that of assigning numerical marks to an individual's examination performance. In both cases there is the initial possibility of a rough and qualitative assessment. An operational procedure is then introduced to turn this into a numerical value. Although this procedure is rather arbitrary, it is nonetheless useful for certain purposes. In the examination case, it helps future employers select their staff. In the probability case, it enables the powerful techniques of mathematical probability to be applied to handling degrees of belief. In both cases, however, the reduction of a complex reality to a single number should perhaps be regarded with some degree of scepticism.

Another point to note[7] is that the method of assigning subjective probabilities has a normative character. The individuals submitting to the process are encouraged to alter their degrees of belief, if necessary, in order to make them coherent. This differentiates subjective probability from a natural sciences case such as the measurement of temperature by thermometers. Liquids do not, as a result of such a measurement process, make a rational decision to adjust their boiling points!

We see then that the problems of evaluating the results of examinations and of evaluating degrees of belief have a great deal in common, and moreover that these common features are likely to appear in other areas of social life. Operationalism provides a good, or at least satisfactory, way of handling such cases, and so is appropriate for the social sciences. I will next argue that the situations dealt with by the natural sciences are sufficiently different to make operationalism no longer appropriate.

The starting point in the natural sciences is the same or similar. Whether humans considered spatial relations, the phenomena of heat and cold, or the size and density of physical bodies, they must have started with rough qualitative assessments. However, to handle these phenomena theories were gradually evolved, according to which these qualitative phenomena could be explained by simple quantitative parameters. Thus for spatial relations, there was Euclidean geometry with its parameters of length (l) and angle (θ). For heat and cold, there were theories which depended on temperature (T). For physical bodies, there was Newtonian mechanics with the parameter mass (m). These theories were tested out and found to work in practice, and they could then be used to design methods for measuring the parameters involved (l, θ, T or m). This procedure is the basis of the non-operationalist account of conceptual innovation in the natural sciences given in Chapter 7. One could say that it is the absence of successful quantitative theories in the social sciences which renders necessary the introduction of operationalist procedures as an alternative way of making the qualitative quantitative.

I will conclude with a final example designed to illustrate this difference between the social and natural sciences. For the social sciences, let us take as an example the analysis of the behaviour of stock markets. Now stock markets present at first glance a completely quantitative appearance. Each share has at a given moment a perfectly definite numerical price, and these prices vary with time in a way which can be precisely represented by lines on computer screens. Further reflection, however, shows that these seemingly precise numbers are determined in a manner not so very different from examination marks, and in a manner which contains much that is arbitrary in it. Admittedly examination marks are fixed by examiners following rules which are laid down by their institution, whereas on the stock market the prices are determined by the decisions to buy or sell of thousands of investors. Yet the decisions of these investors are far from being independent. Investors influence each other, and, at any given moment, a conventional agreement emerges which largely determines the market price. This conventional agreement arises spontaneously rather than through any explicit decision of a controlling body, and yet it has much the same effect as the conventions which guide examiners when assigning marks.

Two of the most successful theorists of stock markets (successful both in theory and practice) have been Keynes (1936:Chapter 12, 147–64) and Soros (1987/94:Chapter 2, 46–68). What is interesting is that these two, unlike many others in the field, have eschewed the use of mathematics and have presented qualitative accounts of how the stock market functions. One might say that on the surface the stock market is quantitative, but that in reality it is qualitative in character and is propelled forward by thousands of investors making decisions under uncertainty based on qualitative considerations.

This situation can be contrasted with the mechanical behaviour of physical bodies. Anyone looking round to observe leaves falling to the ground, hammer blows breaking stones, waterfalls sending spray into the air, etc. would be struck initially by a complex range of qualitative phenomena. While shares come with numbers attached to them, this is not true of any of the phenomena of terrestrial physics. Indeed Aristotle's *Physics* analyses all these phenomena in purely qualitative terms, and his theories were accepted as definitive for many centuries. Yet the rise and triumph of Newtonian mechanics showed that all these phenomena could be reduced to numbers and mathematics. Here the appearance was qualitative, but the reality turned out to be quantitative.

One can sum up it by saying that Pythagoreanism has been a fruitful philosophy for physics, but a misleading philosophy for the social sciences.

Notes

1 Introductory survey of the interpretations: some historical background

1 There are a number of excellent books on the history of probability in this period. My own account is based principally on the following four: Todhunter (1865), David (1962), Hacking (1975) and Daston (1988). The first two deal mainly with the mathematical developments. Todhunter is very comprehensive, while David is more readable and includes English translations of the Pascal–Fermat letters as Appendix 4. Hacking and Daston mention some mathematical questions, but concentrate more on the philosophical side. They naturally disagree on a number of points, one of which will be discussed in Chapter 2.

2 Equations are numbered by chapter, e.g. 3.1, 3.2, etc. For simplicity, not all equations will be numbered.

3 This formula is read as: 'the probability that mod(p - r/n) is less than e tends to 1 as n tends to infinity.' The precise meaning of this formula will be clear to those familiar with mathematical analysis. However, those who have not studied this branch of mathematics can simply understand it as saying that the probability becomes closer and closer to 1 as n becomes larger and larger. The rate at which the probability approaches its limit 1 is known as the speed of the convergence. Figure 1.1 gives a graphical illustration of convergence to a limit as n ? 8.

4 To illustrate the way in which the binomial distribution tends to the normal distribution, it is necessary to use some mathematical transformations. Suppose we are tossing a coin for which Prob(heads) = p, and we obtain r heads in n tosses. Then Z = the relative frequency of heads (r/n) has the binomial distribution

$$\text{Prob}(Z = r/n) = {}^{n}C_{r}\, p^{r}(1 - p)^{n - r}$$

This has mean p and standard deviation

$$\sqrt{\frac{p(1-p)}{n}}$$

It is convenient to consider the standardised variable

$$X = \frac{r/n - p}{\sqrt{\frac{p(1-p)}{n}}}$$

The distribution of X tends to the normal distribution with zero mean ($\mu = 0$) and unit standard deviation ($\sigma = 1$) as $n \to \infty$. To illustrate this we plot the values of X for fixed n and p, and $r = 0$, 1, ..., n on the x-axis, and at each point we plot along the y-axis the value of the binomial distribution multiplied by $\sqrt{np(1 - p)}$. This is the scaled binomial which is compared with the normal distribution with zero mean and unit standard deviation, since the limit theorem states

$$\sqrt{np(1-p)}\,{}^{n}C_{r}p^{r}(1-p)^{n-r} \to \frac{1}{2\pi}\exp\left(-x^{2}/2\right)$$

In Figure 1.1 this procedure was carried out for (a) $p = 0.6$, $n = 5$, and (b) $p = 0.6$, $n = 30$. I am most grateful to my son Mark Gillies for doing the computer graphics. It is noteworthy that for n as small as 30, the approximation of the binomial to the normal distribution is very good. For further mathematical details including two proofs of De Moivre's theorem, one using a modern approach and the other an approach closer to De Moivre's original one, see Cramér (1946: 198–203).

2 The classical theory

1 Chapter 4 contains a full account of the axioms of probability, including an explanation of what is meant by 'finite additivity'.
2 This objection was made to me in conversation by Dr Tony Dale.

3 The logical theory

1 In the last fifteen or so years there has been a great deal of scholarly work on Cambridge in this period, and this has been very helpful for understanding the intellectual currents of the time. For my account in this chapter, I have found the following works very helpful: Bateman (1988, 1996), Davis (1994), Monk (1990, 1996) and Skidelsky (1983, 1992). I have also benefited from reading Childers (1996), which contains useful chapters on the logical theory of probability, both in Keynes's and Carnap's version.
2 For a more detailed comparison between Moore and Keynes as regards Platonism and intuition, see Davis (1994: 10–28).
3 For further details, see Keynes (1921).
4 This example was suggested by a member of the audience when I was lecturing on this topic on one occasion.
5 This claim is argued for in detail in Gillies (1987).
6 The sketch given is rough and designed only to illustrate one of the key features of the argument of Bose and Einstein. For a fuller account with mathematical details see, for example, Born (1935: 268–76).

4 The subjective theory

1 A good discussion of these criticisms of Ramsey's is to be found in Cottrell (1993: 30–2).
2 The heroine and hero of this betting scenario are named after the principal characters in Samuel Richardson's novel of 1740 *Pamela; or, Virtue Rewarded.* Pamela Andrews (Ms A) is a servant girl in the home of Mr B (always referred to thus in the novel). Mr B, who is very rich, attempts to seduce Pamela, but she virtuously refuses his advances, and eventually he decides to marry her. The novel was a best seller at the time of its publication and exerted an enormous influence

on the development of European literature. Presumably in Richardson's fictional setting, it must have been important for Ms A to ascertain Mr B's degrees of belief in various propositions.

3 For an interesting discussion of the money versus utility problem which is more sympathetic to the views of Ramsey, see Sahlin (1990: 41–3).

4 The proof which follows is based on De Finetti 1937, but expanded to fill in the details. A shorter but mathematically more sophisticated proof is to be found in Paris (1994: 19–23).

5 This was pointed out to me by Ladislav Kvasz.

6 A full account of his views on the question of finite versus countable additivity can be found in De Finetti (1970), and I discuss these views in my review of the book (Gillies 1972b: 142–5). In that article I give references to the original Italian edition of De Finetti's book, but in what follows here my references will be to the English translation which appeared in 1974.

7 I learnt of this example from a typescript version of Popper's (1957a), which was circulating in LSE when I was a graduate student there in 1966–8. Popper considers a situation in which the Sun has risen 1,000,000 times in succession but then fails to rise for 10 days. He uses this to criticise the subjective theory of learning in general terms for giving too much authority to past experience, and making a revision of our ideas practically impossible. Although nearly all of the typescript is reprinted in Popper (1957a, 1983), this example is rather curiously omitted. A possible reason is that the example is not effective against all versions of the subjective theory of learning. As Howson and Urbach point out (1989: 81), Bayesianism implies falsificationism in the sense that refuted hypotheses acquire probability 0. Let us consider then a version of subjective Bayesianism which is concerned with the learning of general laws in the sense of trying to assign probabilities to such laws in the light of evidence. Such an approach would have assigned a probability to the universal law that the Sun rises every morning in the light of the 1,000,000 sunrises in succession. However, this probability would drop to zero after the first failure. Thus Popper's example is not a good argument against all versions of the subjective theory of learning, but it does yield a very strong argument against the Rule of Succession as I will show in what follows.

8 The game of red or blue is described in Feller (1950: 67–95), which contains an interesting mathematical analysis of its curious properties. Popper read of the game in Feller, and he had the idea of using it to argue against various theories of induction. Popper (1957a: 358–60) (reprinted 1983: 301–5) uses the game to criticise what he calls 'the simple inductive rule', while later (Popper 1957a: 366–7, reprinted in 1983: 323–4) he uses the game to try to prove the impossibility of an inductive logic. The first of these arguments seems to me valid, and I have adapted it to produce the criticism of De Finetti's exchangeability reduction given here. The second of Popper's arguments seems to me less convincing, since it is perfectly possible that an inductive logic could be devised which could accommodate cases like the game of red or blue. Indeed I give arguments in favour of the possibility of an inductive logic (Gillies 1996: 98–112).

9 The mathematical part of Albert's argument is to be found in Albert (1999), where Theorem 1 is what is here called the *Anything Goes Theorem.* The more philosophical part of the argument will be published soon. I am most grateful to Max Albert for sending me an unpublished typescript with a full discussion of both the mathematical and philosophical sides of the argument, as well as for some helpful discussions of the question and its relation to the argument involving the game of red or blue.

10 An interesting account of W. E. Johnson's contribution to this question is to be found in Zabell (1989).

11 De Finetti initially used the term 'equivalent' (in Italian *equivalente*), but the term 'exchangeable' has now become standard.

5 The frequency theory

1 A critical account of Mach's operational definition of mass is to be found in Gillies (1972a), and also Gillies (1973: Chapter 1, 37–47). The latter reference gives more details about the relationship between Mach and Von Mises.

2 The following proof is based on one suggested to me by Jon Williamson. I am very grateful for his help in this matter.

6 The propensity theory: (I) general survey

1 Popper first presented the propensity theory at a conference in the University of Bristol. However, as he could not attend himself, his paper (Popper 1957b) was read by his then student Paul K. Feyerabend.

2 I regret that shortage of space prevents me from discussing in detail several other important contributions to the field, including Hacking (1965) and Mellor (1971). For a discussion of Mellor's version of the propensity theory, see Salmon (1979).

3 On this topic see Fetzer (1993).

4 This was pointed out to me by David Corfield and Jon Williamson.

5 The reader may be interested to know that Francesca did get her motor scooter and has ridden it about Rome since the mid-1980s without having an accident.

6 This argument was suggested to me by Ladislav Kvasz.

7 This at least has been my experience of coin tossing, but David Miller assures me that there is a mechanical coin-tossing apparatus which is guaranteed to produce heads each time. I have not seen such an apparatus to check the claim myself.

8 This was in fact the position which I adopted in my 1973 book *An Objective Theory of Probability*. The theory developed in that book was objective and non-frequency, but yet I argued against calling it a propensity theory, partly because it differed in some respects from Popper's theory. Indeed I had at that time some general doubts about the use of the term 'propensity' even for Popper's own views (cf. 1973: 149–50). Subsequently, however, the term 'propensity' became well established in the literature, and it has taken on the broader meaning of an objective but non-frequency view of probability. I would therefore now re-classify my earlier position as one particular example of a propensity theory. I had some discussions with Popper on this point after my book had appeared. Interestingly Popper favoured using the term 'propensity' in a general sense rather than as specifically referring to his own views.

9 Obviously this set of axioms would normally be the Kolmogorov axioms, but, as we shall see, there are some versions of the propensity theory, notably Fetzer's, in which propensities do not satisfy Kolmogorov's axioms, but a different set of axioms. In my own approach (see Chapter 7) propensities satisfy the Kolmogorov axioms, but can be further characterised by adding an additional axiom.

10 The distinction between long-run and single-case propensity theories is taken from Fetzer (1988: 123, 125–6). However, I am using the terminology in a slightly different sense from Fetzer. Fetzer takes the 'long run' to refer to infinite sequences, while, as already explained, I am using 'long run' to refer to long, but still finite sequences of repetitions.

7 The propensity theory: (II) development of a particular version

1 Somewhat more extensive accounts of the ideas in this section are to be found in Gillies (1972a, 1973: Part one, 37–74).

2 A detailed historical justification of this claim is to be found in Gillies (1972a: 8–11, or 1973: 48–50).

3 A proof of this result is given in Rutherford (1951: 66–71).
4 This article is reprinted with little alteration in Popper (1983: 132–46).
5 The problems of formulating a falsifying rule for probability statements and of examining its agreement with statistical practice are dealt with in more mathematical detail in Gillies (1971, 1973: Part 3, 161–226). I have given a more brief and informal account here partly in order to keep the mathematics used as simple as possible, but partly also because I want in the present book to focus on the philosophy of probability rather than on the problems of the foundations of statistics – important and fundamental though these are.

The complexities involved in examining the agreement between a proposed FRPS and statistical practice are not, however, just mathematical, because there are disagreements about what should constitute statistical practice. As I remarked in the text the majority of statisticians do implicitly use methodological falsificationism. Howson and Urbach who argue for the alternative Bayesianism approach actually go so far as to suggest that much of standard statistical practice should be abandoned. Thus they write, speaking of Fisher, Neyman, Pearson and others:

> ... it is fair to say that their theories, especially those connected with significance testing and estimation, which comprise the bulk of so-called classical methods of statistical inference, have achieved pre-eminence in the field. The procedures they recommended for the design of experiments and the analysis of data have become the standards of correctness with many scientists.
>
> In the ensuing chapters we shall show that these classical methods are really quite unsuccessful, despite their influence amongst philosophers and scientists, and that their pre-eminence is undeserved.
>
> (1989: 11)

True to their word, Howson and Urbach in three later chapters of their book make extensive criticisms of the standard theory of statistical testing (1989: Chapters 5–7, 121–76). I devote a substantial part of my review of their book (Gillies 1990: 90–8) to trying to answer these objections, and to arguing that it is unlikely that the standard methods of statistical testing will be given up.

Albert (1992) is a most important recent contribution to the problem of falsificationism and statistical inference. Among other things, it contains some valuable remarks about the chi-square test from this point of view.

8 Intersubjective probability and pluralist views of probability

1 This proof was suggested to me by Professor D. V. Lindley (private correspondence).
2 For applications to economics, see Gillies (1988b) and Gillies and Ietto-Gillies (1991). For applications to the confirmation of scientific hypotheses, see Gillies (1991).
3 This important point was made to me by Ladislav Kvasz in a the course of a long and most beneficial discussion we had on the question of the spectrum from subjective to objective. Ladislav Kvasz also supplied me with the illustrations for Figures 8.1 and 8.2.

9 An example of pluralism: differences between the natural and social sciences

1 The comparison between an economy consisting of millions of people and a gas consisting of millions of molecules is discussed in Farjoun and Machover (1983) particularly Chapter 2, ‘A Paradigm: Statistical Mechanics.’ The point of view of Farjoun and Machover is different, however, from the one expressed here. As the title of their Chapter 2 suggests, they regard the physics case (statistical mechanics) as a suitable model for studying the economic system.

Moreover, they do actually develop their economic theory by analogy with statistical mechanics. I will argue, however, that the cases are different, and, in particular, that they involve different notions of probability.

2 The case in which there is some molecular interaction will be considered later (p. 191).

3 An argument of this kind is used by Farjoun and Machover who write:

> In fact, the development of statistical mechanics has shown that the macroscopic behaviour of such a system depends surprisingly little – much less than envisaged even by Maxwell and Boltzmann – on the precise nature of the microscopic interactions of its particles, but more on the very fact that the system itself is made up of a very large number of constituent parts and, microscopically speaking, has a very large number of 'degrees of freedom'.
>
> (1983: 55–6)

4 A similar case was suggested to me in conversation by Moshé Machover.

5 In particular, medical expert systems are an area in which objective probabilities might be used. For an application of the propensity theory of Chapter 7 to such a system, see Sucar *et al.* (1993).

6 The first edition of Soros's book *The Alchemy of Finance* was published in 1987, and the second edition in 1994. It is accordingly given in the references as Soros 1987/94. In the text quotations from the original edition will be given as Soros 1987, but quotations from the new preface added in 1994 will be given as Soros 1994. This convention makes for clarity, because, as we shall see, Soros's comments in the new preface on the original edition of his book, and modifies his previous opinions on a number of points.

7 I owe this point to Hasok Chang.

References

In general, works are cited by their date of first publication, but the exact edition from which quotations are taken is also specified, and its date given if different from the first edition. Occasionally two places are given where a particular work can be found. In this case the quotations in the text are from the first cited place, but the second is added as it might be more accessible to the reader. English translations of works in foreign languages are cited whenever possible.

Adams, E. (1964) 'On Rational Betting Systems', *Archiv für Mathematische Logik und Grundlagenforschung* 6: 7–29, 112–28.

Albert, M. (1992) 'Die Falsifikation Statistischer Hypothesen', *Journal for General Philosophy of Science* 23: 1–32.

—— (1999) 'Bayesian Learning when Chaos looms large', *Economics Letters* 65: 1–7.

Ayer, A. J. (1963) 'Two Notes on Probability', in *The Concept of a Person and other Essays,* Macmillan, 188–208.

Bacon, F. (1620) *Novum Organum,* English translation in R. L. Ellis and J. Spedding (eds), *The Philosophical Works of Francis Bacon,* Routledge, 1905, 212–387.

Bateman, B. W. (1988) 'G. E. Moore and J. M. Keynes: A Missing Chapter in the History of the Expected Utility Model', *American Economic Review* 78: 1098–106.

—— (1996) *Keynes's Uncertain Revolution,* University of Michigan Press.

Bayes, T. and Price, R. (1763) 'An Essay towards Solving a Problem in the Doctrine of Chances', reprinted in E. S. Pearson and M. G. Kendall (eds.) *Studies in the History of Statistics and Probability,* Griffin, 1970, 134–53.

Born, M. (1935) *Atomic Physics,* Blackie, 7th Edition, 1966.

Cantelli, F. P. (1935) 'Consideration sur la convergence dans le calcul des probabilités', *Annales de l'institut Henri Poincaré* V: 1–50.

Carnap, R. (1950) *Logical Foundations of Probability,* 2nd Edition, University of Chicago, 1963.

Childers, T. (1996) 'On the Relation between the Normative and the Empirical in the Philosophy of Science', unpublished PhD thesis, University of London.

Church, A. (1936) 'An unsolvable problem of elementary number theory', *American Journal of Mathematics* 58: 345–63.

—— (1940) 'On the concept of a random sequence', *Bulletin of the American Mathematical Society* 46: 130–5.

Cottrell, A. (1993) ‘Keynes’s Theory of Probability and its Relevance to his Economics. Three Theses’, *Economics and Philosophy* 9: 25–51.

Cox, D. R. and Miller, H. D. (1965) *The Theory of Stochastic Processes,* Methuen.

Cramér, H. (1946) *Mathematical Methods of Statistics,* Princeton University Press, 1961.

Czuber, E. (1903/1908–10) *Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlersausgleichung, Statistik und Lebensversicherung.* Teubner.

Daston, L. (1988) *Classical Probability in the Enlightenment,* Princeton University Press.

David, F. N. (1962) *Games, Gods and Gambling. The Origins and History of Probability and Statistical Ideas from the Earliest Times to the Newtonian Era,* Harper.

Davis, J. B. (1994) *Keynes’s Philosophical Development,* Cambridge University Press.

De Finetti, B. (1930a) ‘Fondamenti Logici del Ragionamento Probabilistico’, *Bollettino dell’ Unione Matematica Italiana* 5: 1–3. Reprinted in De Finetti, 1981a, 261–3.

—— (1930b) ‘Funzione Caratteristica di un Fenomeno Aleatorio’, *Memorie della Reale Accademia dei Lincei* IV, 5: 86–133. Reprinted in De Finetti, 1981a, 267–315,

—— (1930c) ‘Problemi Determinati e Indeterminati nel Calcolo della Probabilità’, *Rendiconti della Reale Accademia Nazionale dei Lincei* XII, 9: 367–73. Reprinted in De Finetti, 1981a, 327–33.

—— (1931a) ‘Probabilism’, English translation in *Erkenntnis* 1989, 31: 169–223.

—— (1931b) ‘On the Subjective Meaning of Probability’, English translation in De Finetti, 1993, 291–321.

—— (1936) ‘Statistics and Probability in R. Von Mises’ Conception’, English translation in De Finetti, 1993, 353–64.

—— (1937) ‘Foresight: Its Logical Laws, Its Subjective Sources’, English translation in H. E. Kyburg and H. E. Smokler (eds.), *Studies in Subjective Probability,* Wiley, 1964, 93–158.

—— (1938) ‘Cambridge Probability Theorists’, English translation in *The Manchester School of Economic and Social Studies* 1985, 53: 348–63.

—— (1970) *Theory of Probability.* vols 1 and 2, English translation. Wiley. 1974.

—— (1981a) *Scritti (1926–1930),* Cedam: Padua.

—— (1981b) ‘The Role of “Dutch Books” and “Proper Scoring Rules”’, *British Journal for the Philosophy of Science* 32: 55–6.

—— (1991) *Scritti (1931–1936),* Pitagora: Bologna.

—— (1993) *Induction and Probability,* Clueb: Bologna.

—— (1995) *Filosofia della Probabilità,* Il Saggiatore.

Dostoyevsky, F. M. (1866) *The Gambler,* English translation in Penguin Classics, 1976.

Duhem, P. (1904–5) *The Aim and Structure of Physical Theory,* English translation by Philip P. Wiener of the second French edition of 1914, Atheneum, 1962.

Earman, J. and Salmon, W. C. (1992) ‘The Confirmation of Scientific Hypotheses’, in M. H. Salmon *et al.* (eds), *Introduction to Philosophy of Science.* Prentice Hall, Chapter 2, 42–103.

Edwards, A. W. F. (1978) ‘Commentary on the Arguments of Thomas Bayes’, *Scandinavian Journal of Statistics* 5: 116–18.

Farjoun, E. and Machover, M. (1983) *Laws of Chaos,* Verso.

Feller, W. (1950) *Introduction to Probability Theory and Its Applications,* Third edition, 1971, Wiley.

Fetzer, J. H. (1981) *Scientific Knowledge: Causation, Explanation, and Corroboration,* Boston Studies in the Philosophy of Science, Vol. 69, D. Reidel, Dordrecht.

—— (1982) ‘Probabilistic Explanations’, *PSA* 2: 194–207.

—— (1988) ‘Probabilistic Metaphysics’, in J. H. Fetzer (ed.) *Probability and Causality,* D. Reidel, 109–132.

—— (1991) ‘Critical Notice: Philip Kitcher and Wesley C. Salmon, (eds.), *Scientific Explanation*; and Wesley C. Salmon, *Four Decades of Scientific Explanation. Philosophy of Science’, Philosophy of Science* 58: 288–306.

—— (1993) ‘Peirce and Propensities’, in Edward C. Moore (ed.) *Charles S. Peirce and the Philosophy of Science,* University of Alabama Press, 61–71.

Fisher, R. A. (1925) *Statistical Methods for Research Workers,* Oliver and Boyd.

—— (1935) *The Design of Experiments,* Oliver and Boyd.

Fleck, L. (1935) *Genesis and Development of a Scientific Fact,* English translation, University of Chicago Press, 1981.

Fraser, R. G. F. (1931) *Molecular Rays,* Cambridge University Press.

Fry, T. C. (1928) *Probability and its Engineering Uses,* Van Nostrand.

Galavotti, M. C. (1989) ‘Anti-Realism in the Philosophy of Probability: Bruno de Finetti’s Subjectivism’, *Erkenntnis* 31: 239–61.

—— (1991) ‘The notion of subjective probability in the work of Ramsey and de Finetti’, *Theoria* LXII(3): 239–59.

—— (1994) ‘F. P. Ramsey and the Notion of “Chance”’, *Proceedings of the 17th International Wittgenstein-Symposium (Kirchberg am Wechsel, 14-21 August 1994)* 330–40.

—— (1995) ‘Operationism, Probability and Quantum Mechanics’, *Foundations of Science* 1: 99–118.

—— (1999) ‘Some Remarks on Objective Chance (F. P. Ramsey, K. R. Popper and N. R. Campbell)’, in M. L. Dalla Chiara *et al.* (eds.) *Language, Quantum, Music,* Kluwer, 73–82.

Gibbon, E. (1776–88) *Decline and Fall of the Roman Empire,* John Murray. 1862.

Gillies, D. A. (1971) ‘A Falsifying Rule for Probability Statements’, *British Journal for the Philosophy of Science* 22: 231–61.

—— (1972a) ‘Operationalism’, *Synthese* 25: 1–24.

—— (1972b) ‘The Subjective Theory of Probability’, *British Journal for the Philosophy of Science* 29: 138–57.

—— (1973) *An Objective Theory of Probability,* Methuen.

—— (1982) *Frege, Dedekind, and Peano on the Foundations of Arithmetic,* Van Gorcum.

—— (1987) ‘Was Bayes a Bayesian?’, *Historia Mathematica* 14: 325–46.

—— (1988a) ‘Induction and Probability’, in G. H. R. Parkinson (ed.), *An Encyclopaedia of Philosophy,* Chapter 9, 179–204.

—— (1988b) ‘Keynes as a Methodologist’, *British Journal for the Philosophy of Science* 39: 117–29.

—— (1990) ‘Bayesianism versus Falsificationism. Review of Howson and Urbach 1989’, *Ratio* New Series III(1): 82–98.

—— (1991) ‘Intersubjective Probability and Confirmation Theory’, *British Journal for the Philosophy of Science* 42: 513–33.

—— (1996) *Artificial Intelligence and Scientific Method,* Oxford University Press.

Gillies, D. A. and Ietto-Gillies, G. (1991) ‘Intersubjective Probability and Economics’, *Review of Political Economy* 3(4): 393–417.

Gnedenko, B. V. (1950) *The Theory of Probability,* English translation, Chelsea, 1962.

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

—— (1975) *The Emergence of Probability. A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference,* Cambridge University Press, Paperback edition, 1984.

Hicks, J. (1979) *Causality in Economics,* Blackwell.

Hilbert, D. (1899) *Foundations of Geometry,* English translation, Open Court, 1971.

Howson, C. and Urbach, P. (1989) *Scientific Reasoning. The Bayesian Approach,* Open Court.

Humphreys, P. (1985) 'Why Propensities cannot be Probabilities', *The Philosophical Review* 94: 557–70.

Iversen, G. R., Longcor, W. H., Mosteller, F., Gilbert, J. P., and Youtz, C. (1971) 'Bias and Runs in Dice Throwing and Recording: a Few Million Throws', *Psychometrika* 36(1): 1–19.

Jaynes, E. T. (1973) 'The Well-Posed Problem', *Foundations of Physics* 4(3): 477–92.

Jeffreys, H. (1939) *Theory of Probability,* Oxford University Press.

Kendall, M. G. and Babington Smith, B. (1938) 'Randomness and Random Sampling Numbers', *Journal of the Royal Statistical Society* 101: 147.

Kendall, M. G. and Babington Smith, B. (1939a) 'Randomness and Random Sampling Numbers', *Journal of the Royal Statistical Society (Supplement)* 6: 57.

Kendall, M. G. and Babington Smith, B. (1939b) *Tables of Random Sampling Numbers,* Cambridge University Press.

Keynes, J. M. (1921) *A Treatise on Probability,* Macmillan, 1963.

—— (1936) *The General Theory of Employment, Interest and Money.* The Collected Writings of John Maynard Keynes, Vol. VII, Macmillan, Cambridge University Press for the Royal Economic Society, 1993.

—— (1938) 'My Early Beliefs', in The Collected Writings of John Maynard Keynes, Vol. X, *Essays in Biography,* Macmillan, Cambridge University Press for the Royal Economic Society, 1985.

Kolmogorov, A. N. (1933) *Foundations of the Theory of Probability,* Second English Edition, Chelsea, 1956.

Kuhn, T. S. (1962) *The Structure of Scientific Revolutions,* University of Chicago Press.

Lad, F. (1983) 'The Construction of Probability Theory', *Science and Society* 47: 285–99.

Lad, F. (1996) *Operational Subjective Statistical Methods,* Wiley.

Lakatos, I. (1968) 'Changes in the Problem of Inductive Logic', in *Philosophical Papers,* Vol. 2, Cambridge University Press, Paperback edition, 1983, Chapter 8, 128–200.

Laplace, P. S. (1814) *A Philosophical Essay on Probabilities,* English Translation of the 6th French Edition, Dover, 1951.

Locke, J. (1690) *An Essay concerning Human Understanding,* Everyman Library, J. M. Dent and Sons, 1961.

McCurdy, C. S. I. (1996) 'Humphreys's Paradox and the Interpretation of Inverse Conditional Propensities', *Synthese* 108: 105–25.

Mach, E. (1883) *The Science of Mechanics: A Critical and Historical Account of its Development,* 6th American Edition, Open Court, 1960.

Mellor, D. H. (1971) *The Matter of Chance,* Cambridge University Press.

Miller, D. W. (1994) *Critical Rationalism. A Restatement and Defence,* Open Court.

—— (1996) 'Propensities and Indeterminism', in A. O'Hear (ed.) *Karl Popper: Philosophy and Problems,* Cambridge University Press, 121–47.

Milne, P. (1986) 'Can there be a Realist Single-Case Interpretation of Probability?', *Erkenntnis* 25: 129–32.

Monk. R. (1990) *Ludwig Wittgenstein. The Duty of Genius,* Jonathan Cape.

—— (1996) *Bertrand Russell. The Spirit of Solitude,* Jonathan Cape.

Newton, I. (1687) *Philosophiae Naturalis Principia Mathematica,* Andrew Motte's English translation of 1729, revised by Florian Cajori, University of California Press, 1960.

Paris, J. B. (1994) *The Uncertain Reasoner's Companion. A Mathematical Perspective,* Cambridge University Press.

Pascal, B. (1670) *Pensées,* English translation, Penguin, 1972.

Pearson, K. (1900) 'On the Criterion that a given System of Deviations from the Probable in the case of a Correlated System of Variables is such that it can be reasonably supposed to have arisen from Random Sampling', Reprinted in *Karl Pearson's Early Statistical Papers,* Cambridge University Press, 1956, 339–57.

Peirce, C. S. (1910) 'Notes on the Doctrine of Chances', Reprinted in *Essays in the Philosophy of Science,* The American Heritage Series, Bobbs-Merrill, 1957, 74–84.

Poincaré, H. (1902) *Science and Hypothesis,* English translation, Dover, 1952.

Popper, K. R. (1934) *The Logic of Scientific Discovery,* 6th revised impression of the 1959 English translation, Hutchinson, 1972.

—— (1957a) 'Probability Magic or Knowledge out of Ignorance', *Dialectica* 11(3/4): 354–74.

—— (1957b) 'The Propensity Interpretation of the Calculus of Probability, and the Quantum Theory', in S. Körner (ed.) *Observation and Interpretation, Proceedings of the Ninth Symposium of the Colston Research Society, University of Bristol,* 65–70 and 88–9.

—— (1957c) 'The Aim of Science', *Ratio* 1(1): 24–35.

—— (1959a) *New Appendices to the Logic of Scientific Discovery,* in 6th revised impression of the 1959 English translation, Hutchinson, 1972, 307–464.

—— (1959b) 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science* 10: 25–42.

—— (1963) *Conjectures and Refutations,* Routledge and Kegan Paul.

—— (1983) *Realism and the Aim of Science,* Hutchinson.

—— (1990) *A World of Propensities,* Thoemmes.

Ramsey, F. P. (1926) 'Truth and Probability', in Ramsey, 1931, 156–98. Reprinted in H. E. Kyburg and H. E. Smokler (eds), *Studies in Subjective Probability,* Wiley, 1964, 61–92.

—— (1928) 'Further Considerations', in Ramsey, 1931, 199–211.

—— (1929) 'Last Papers. C. Probability and Partial Belief', in Ramsey, 1931, 256–7.

—— (1931) *The Foundations of Mathematics and other Logical Essays,* Routledge and Kegan Paul.

—— (1991) *Notes on Philosophy, Probability and Mathematics,* Bibliopolis.

Reichenbach, H. (1949) *The Theory of Probability,* University of California Press.

Richardson, S. (1740) *Pamela; or Virtue Rewarded,* Penguin Classics, 1985.

Runde, J. (1994) 'Keynes After Ramsey: In Defence of *A Treatise on Probability', Studies in History and Philosophy of Science* 25(1): 97–121.

—— (1996) 'On Popper, Probabilities and Propensities', *Review of Social Economy,* LIV(4): 465–85.

Russell, B. (1905) 'On Denoting', in R. C. Marsh (ed.) *Bertrand Russell. Logic and Knowledge. Essays 1901–50,* Allen and Unwin, 1956, 41–56.

—— (1912) *The Problems of Philosophy,* Williams and Norgate, n.d.

—— (1914) *Our Knowledge of the External World,* Allen and Unwin, 1961.

—— (1959) *My Philosophical Development,* Allen and Unwin, 1969.

—— (1967) *The Autobiography of Bertrand Russell,* Vol. 1. *1872–1914,* Allen and Unwin.

Rutherford, D. E. (1951) *Classical Mechanics,* 2nd Edition, Oliver and Boyd, 1957.

Ryder, J. M. (1981) 'Consequences of a Simple Extension of the Dutch Book Argument', *British Journal for the Philosophy of Science* 32: 164–7.

Sahlin, N. -E. (1990) *The Philosophy of F. P. Ramsey,* Cambridge University Press.

Salmon, W. C. (1979) 'Propensities: a Discussion Review of D. H. Mellor *The Matter of Chance', Erkenntnis* 14: 183–216.

Sambursky, S. (1954) *The Physical World of the Greeks,* English paperback edition, Routledge and Kegan Paul, 1963.

Skidelsky, R. (1983) *John Maynard Keynes.* Vol. 1. *Hopes Betrayed 1883–1920,* Macmillan.

—— (1992) *John Maynard Keynes.* Vol. 2. *The Economist as Saviour 1920–1937,* Macmillan.

Soros, G. (1987/94) *The Alchemy of Finance. Reading the Mind of the Market,* 2nd Edition, 1994, John Wiley.

Soros, G. (1992) Introduction to W. Newton-Smith and J. Tianji (eds.) *Popper in China,* Routledge.

Student (Gosset, W. S.) (1908) 'The probable error of a mean', *Biometrika* 6: 1–25.

Sucar, L. E., Gillies, D. F., and Gillies, D. A. (1993) 'Objective Probabilities in Expert Systems', *Artificial Intelligence* 61: 187–208.

Todhunter, I. (1865) *A History of the Mathematical Theory of Probability from the time of Pascal to that of Laplace,* Chelsea, 1965.

Von Mises, R. (1919) 'Grundlagen der Wahrscheinlichkeitsrechnung', Reprinted in Von Mises, 1964b, 57–106.

—— (1928) *Probability, Statistics and Truth,* 2nd revised English edition, Allen and Unwin, 1961.

—— (1938) 'Ernst Mach und die empiristische Wissenschaftsauffassung', Reprinted in Von Mises, 1964b, 495–523.

—— (1940) 'Scientific Conception of the World: On a New Textbook of Positivism', Reprinted in Von Mises, 1964b, 524–9.

—— (1964a) *Mathematical Theory of Probability and Statistics,* Academic Press.

—— (1964b) *Selecta II,* American Mathematical Society.

Wald, A. (1937) 'Die Widerspruchsfreiheit des Kollektivsbegriffes', *Ergebnisse eines Mathematischen Kolloquiums* 8: 38–72.

—— (1938) 'Die Widerspruchsfreiheit des Kollektivbegriffes', in *Wald: Selected Papers in Statistics and Probability,* MacGraw-Hill, 1955, 25–41.

Williamson, J. O. D. (1999) 'Countable Additivity and Subjective Probability', *British Journal for the Philosophy of Science,* 50: 401–416.

Wittgenstein, L. (1921) *Tractatus Logico-Philosophicus,* Routledge and Kegan Paul. 1963.

Zabell, S. L. (1989) 'The Rule of Succession', *Erkenntnis* 31: 283–321.

Index