Some Notes on the Methodology of Actuarial Science

Edinburgh, UK.

7th April 2020.

Table of Contents

Introduction

"Philosophy, though unable to tell us with certainty what is the true answer to the doubts which it raises, is able to suggest many possibilities which enlarge our thoughts and free them from the tyranny of custom." *Bertrand Russell, The Problems of Philosophy, 1912.*

This famous Bertrand Russell quotation is representative of the spirit with which this work has been undertaken. This work aims to consider the methodology of actuarial science through the lens of the philosophy of science. In doing so, few 'true answers' will be irrefutably established. But I hope the work can offer some observations and suggest some possibilities that stimulate actuaries to consider different approaches to actuarial science and modelling.

The following pages seek to step back from our day-to-day actuarial methods by questioning the robustness of their premises and logic, identifying their implied limits, and speculating on where alternative methods may offer greater prospects of success (and, indeed, sometimes re-considering what actuarial success might look like). Over the course of the work some specific and detailed answers to the methodological questions raised will be offered. But these answers can only be regarded as tentative and fallible opinions, rather than claims to truth. In the spirit of Russell above, the value of the work is not necessarily found in determining the right answers, but in demonstrating the benefit of asking the questions. I hope these pages stimulate more actuaries to ask these methodological questions and share their own tentative answers.

Relatively little of this type of study has been published by actuaries, especially in the 21st century. The experience of undertaking this study allows me to testify to some of the good reasons for that. The process necessarily involves considering professional practice areas outside the methodologist's own specific area of expertise and offering some constructive but contentious advice. It requires a willingness to stick one's head above the parapet far enough to scan the horizon without worrying too much about being shot in the eye. This leads me to vouch for another quotation, this time from a methodologist in economics:

"The study of methodology is an agonizing task; writing a book on the subject requires the skills of an individual who is at once presumptuous and masochistic."[1]

In truth, undertaking this work has been far from agony…in fact, this has been the most stimulating and enjoyable body of actuarial study that I have ever undertaken. But I often identified with the purported character traits.

The Structure of this Work and How to Read It

So, the following chapters are intended to free our actuarial minds from the tyranny of custom by examining the doubts raised by a philosophically-minded review of the methodology of actuarial science. This is performed in two stages. First, Part One provides an overview and discussion of some of the key areas of the philosophy of science that might be relevant to the methodology of actuarial science. Little of this discussion is specific to actuarial science, but it is written with actuarial contexts in mind where possible. Part Two then seeks to apply the ideas and insights developed in Part One to the specifics of actuarial science and its methodology.

[1] Caldwell (1994), p. 1.

Part One is quite long and some of it is doubtless less than strictly relevant to actuarial science. The less philosophically curious or, indeed, the already philosophically well-read reader may choose to regard Part One as a sort of extended appendix - and thereby avoid reading it except where Part Two suggests it is useful. These readers can skip straight on to Part Two, which has been written with this approach in mind, and which therefore has frequent referencing and signposting of the relevant Part One material.

Part One: Probability and Methodology in Science

1. Philosophical theories of probability

What do we mean by the word 'probability'? At first sight, the answer may appear so intuitive and obvious as to make the question seem frivolous. Isn't it merely the likelihood of some well-defined event happening over some well-defined time interval or amongst some specified number of instances? And didn't the great Pierre Simon Laplace adequately address the question 200 years ago in the very first sentence of his treatment of probability:

"The probability for an event is the ratio of the number of cases favourable to it, to the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible."[2]

By the end of this chapter the reader may have a fuller appreciation of the different interpretations that can be placed on Laplace's definition. These alternative interpretations will demonstrate that the word 'probability' can mean surprisingly different things to different people. This is, in part, simply a consequence of the sheer breadth of linguistic uses to which the word is often put. But we shall find that these different interpretations are not merely linguistic – each approach comes with its own premises, logical development, mathematical theory and accompanying philosophical (epistemological) outlook.

A few examples will help to illustrate the diversity of circumstances which a definition of probability might be required to encompass: what is the probability that the next throw of an unbiased die will produce a 6? If my first six throws of an unknown die produce 6 6s, what is the probability that the die is unbiased? What is the probability that I catch a fish on my next fishing trip? What is the probability that the Big Bang theory of the universe is correct? By how much did this probability increase when cosmic microwave background radiation was discovered in 1964? These examples highlight that the notion of probability can arise across very wide-ranging sets of circumstances. Just as importantly, small differences in the specific context in which probability is applied can serve as a potentially great source of confusion. Indeed, the subtleties that can arise in attempting a definition of probability have been the source of philosophical and scientific debate for at least the last few centuries.

Before proceeding to further consider these subtleties, it may be helpful to note a couple of fundamental distinctions that will recur when discussing the meaning and usage of probabilities. First, it is helpful to always bear in mind that probabilistic thinking can be applied in two 'different directions'. In one direction, we begin with a population of subjects or events whose probabilistic characteristics are *assumed to be known* (for example, suppose we can take it as known that a given die is unbiased). In this setting, we may wish to derive the *properties of a sample* from that population, which is often called the sampling distribution (for example, the probability distribution of the sum of the next ten rolls of the dice). In this setting, probability is 'merely' a mathematical discipline: the sample properties are mathematically deduced from the premises that are assumed to be known (and this will include assumptions about the sampling procedure as well as the properties of the population. For example, we may assume the die is rolled in a way that makes it equally likely that any side lands face up).

In the other direction, our starting point is a set of *observations from a sample of population data* (in the above example, the results of each of the next ten rolls of the die). We then use this sample data

[2] Laplace (1812)

to make an *inference* about the characteristics of a population (to continue the example, this could be: what is the probability the die will land on a 6; or, is there sufficient evidence to reject the hypothesis that the die is unbiased). Probabilities inferred when working in this second direction have historically been referred to as indirect or inverse probabilities. Today, we would typically refer to this form of activity as statistics or statistical inference. Whatever we call it, it inevitably involves greater conceptual and philosophical challenges than the mathematical manipulation of known probabilities. Whilst the notion of probability has been recognised in some form for thousands of years and was explicitly mathematically developed in the mid-17th century, no progress was made in establishing a scientific basis for statistical inference until the pioneering work of philosophically-minded mathematicians such as Bayes and Laplace at the end of the 18th century. And it has been steeped in controversy ever since.

Moving on to the second key distinction we wish to make before discussing probability definitions in fuller detail. Since the development of probabilistic thinking in the 17th and 18th centuries, there has been some recognition of two quite different ways in which the notion of probability can arise – what Ian Hacking, the philosopher and historian of probability and statistics, has referred to as the 'Janus-faced' character of probability[3]. One face considers probability in the context of the frequency with which a particular kind of thing occurs amongst a *large group of similar things* (this is sometimes referred to as an objective probability, or, unhelpfully, as a statistical probability; Hacking used the term 'aleatory' to refer to probability in this setting, but his terminology hasn't really caught on); but probability is also widely used as a measure of how strongly we believe a *specific single thing* to be true (such as the Big Bang theory). This latter use of probability is closely related to the philosophical discipline of epistemology – the theory of knowledge. Does knowledge only arise when we know something is true? What if we do not have a sufficiently complete understanding of the world to claim we *know* something is true, but we have reason to think it is *probably* true? What does 'probably' mean in this setting? Can probability theory be useful employed to help deal with such forms of knowledge?

Below we consider how philosophers and probabilists have sought to answer some of these questions in competing theories of what is meant by 'probability'.

1.1 Theories of objective probability

We begin with theories of objective probability. This is the natural place to start as objective theory is (arguably) the most straightforward and intuitive way of thinking about probability. It is also the *narrowest* way of thinking about probability – as we shall see, objective theory's answer to some of the more philosophically-demanding probabilistic questions is essentially that what is being discussed is not probability and is therefore no concern of theirs.

To an objectivist, probability is a real property of a physical phenomenon that is, at least in principle, capable of objective measurement. In the words of Sir Karl Popper, a leading 20th century proponent of the objective approach to probability, probabilities are 'abstract relational properties of the physical situation, like Newtonian forces'[4].

The 'standard' objective theory of probability is the *frequentist* (or frequency) theory. It has been argued that Aristotle was the first to allude to a frequency theory of probability when he wrote in *Rhetoric* that probability is 'that which happens generally but not invariably'[5]. But the pioneers of

[3] Hacking (1975)
[4] Popper (1959), p. 28.
[5] Nagel (1939), p. 18.

probability of the 18th and 19th centuries such as Laplace, Bayes and Poisson did not (explicitly) define probability in a frequentist way. It wasn't until the mid-19th century that the frequency theory started to be developed in an explicit and mathematical form. The best-known exposition of the frequency theory from that era is provided by the Cambridge philosopher John Venn in his book *Logic of Chance*, which was first published in 1866[6]. The classic 20th century treatment of the frequency theory was produced by the German mathematician and physicist Richard Von Mises in 1928[7]. We shall focus mainly on Von Mises' theory as it is more mathematically rigorous than Venn's and is generally regarded as superseding it. We will still, however, have occasion to reflect on some of the ideas and perspectives that Venn articulated in his early development of the theory.

At the core of Von Mises' frequency probability theory is his mathematical concept of a *collective*, which he defines as 'a sequence of uniform events or processes which differ by certain observable attributes'[8]. For example, one of these 'uniform events' could be the throw of a die and the observable attribute could be the die landing 1. Von Mises defines the probability of an attribute *A* as:

$$P(A) = \lim_{n \to \infty} \frac{m(a,n)}{n}$$

where *n* is the number of events in the collective; *A* is some specified attribute of the events in the collective; and *m(a,n)* is the number of events with attribute *A* in a collective of size *n*.

The idea of defining probability by reference to a set of events when the number of events in the set tends to infinity was first introduced by Venn and is sometimes referred to as the *Venn limit*. In order to ensure the probability as defined above is mathematically well-behaved, Von Mises had to postulate two axioms: an axiom of convergence (which states that the mathematical limit used in the definition exists); and an axiom of randomness (which specifies that the order of the elements in the collective is not arranged in any determinable way).

The Venn limit and Von Mises' axioms are discussed further below. But before we consider the logical and philosophical issues that arise with this definition, it is worth noting that Von Mises' probability definition seems to fit well with the essence of what is often meant by 'probability'. It is clearly different from, but not obviously inconsistent with, the definition given by Laplace, quoted above. Like the Laplace definition, it is based on a ratio of 'favourable' cases to the total number of 'equally possible' cases. Bernoulli's law of large numbers holds true when probability is defined using the Von Mises definition. Indeed, the law of large numbers becomes tautological under Von Mises' definition of probability. Some philosophers of science and probability such as Sir Harold Jeffreys have argued that this suggests Bernoulli had a much broader definition of probability in mind when he derived the law of large numbers (otherwise his derivation would have been a lot simpler.)[9] We will return to this issue of the narrowness of the frequentist definition of probability later. First, we briefly consider the three key philosophical issues that arise with Von Mises' definition.

Philosophical Concerns with Probability as a Limiting Frequency

Some philosophers have objected to a definition of probability that relies on the idea of a collective or set of elements whose size tends towards infinity. Sir Harold Jeffreys, a Cambridge Professor of geophysics and astronomy, was an important 20th century thinker and writer on probability and its

---

6 Venn (1888)
7 Von Mises (1928)
8 Von Mises (1928), p. 12.
9 Jeffreys (1961), p. 404.

philosophical challenges whose work we will discuss further later. He rejected the frequency theory of probability, and the Venn limit was one of his bones of contention. He argued that the need for *a priori* axioms such as those used by Von Mises in his frequency definition defeated the purpose of so explicitly defining the probability. To Jeffreys, it would be simpler and no less ambiguous to simply take the probability as a basic concept that does not have an explicit definition but is simply defined by the rules that govern its behaviour:

"The very existence of the probability on Venn's [or Von Mises'] definition requires an *a priori* restriction on the possible orders of occurrence [an axiom of randomness in Von Mises' terminology]. No supporter of this view has succeeded in stating the nature of this restriction, and even if it were done it would constitute an *a priori* postulate, so that this view [a frequency theory definition of probability] involves no reduction of the number of postulates involved in the treatment of probability as an undefined concept with laws of its own."[10]

This view is not, however, universally shared today. The mathematical rigour of Von Mises' axiom of randomness was contested by Jeffreys and others. But the mathematician Alfonso Church[11] provided (a few years after Jeffreys' comments above were published) a refinement of Von Mises' definition of randomness which many have argued furnish it with ample mathematical rigour[12].

The concerns with Von Mises' axiom of randomness are concerns with the frequency definition's abstract logical properties. Jeffrey's above objection to the Venn limit is also based on logical grounds. But concerns have also been raised about the empirical interpretation and implementability that arises with a definition based on the Venn limit. Quoting Jeffreys once more: "A definition is useless unless the thing defined can be recognized in terms of the definition when it occurs. The existence of a thing or the estimate of a quantity must not involve an impossible experiment."[13] Jeffreys therefore ruled out "any definition of probability that attempts to define probability in terms of infinite sets of observations, for we cannot in practice make an infinite number of observations".[14]

This objection can be viewed as what a philosopher of science may call an *operationalist* perspective. Operationalism is a philosophical doctrine of science that argues that we cannot know the meaning of something unless we have a way of empirically measuring it by means of a set of well-defined empirical operations (which would typically be taken to mean experimental operations or some other form of controlled observation). It therefore requires the basic objects of scientific theories to be defined with reference to some observable phenomenon (and so is opposed to the use of theoretical or non-observable entities). It was most notably developed as a philosophical doctrine by the experimental physicist and Nobel laureate, Percy Williams Bridgman in his 1927 book, *The Logic of Modern Physics*. Operationalism was a popular perspective amongst philosophers of science during the second quarter of the twentieth century (when Jeffreys was in his prime). It was in sympathy with the empiricist *zeitgeist* of the era, and found some support amongst the logical positivists of the Vienna Circle (who will be discussed further in Chapter 2).

Operationalism was never universally accepted, however (Einstein, and other theoretical physicists were, perhaps unsurprisingly, not in favour). It has retained some support ever since its 1930s heyday, but it has not necessarily been an orthodox perspective since the start of the post-war era.

[10] Jeffreys (1937), p. 221.
[11] Church (1940)
[12] See Gillies (2000), p. 105-9 for a fuller discussion.
[13] Jeffreys (1961), p. 8.
[14] Jeffreys (1961), p. 11.

The arguments in support of the importance of the role of non-observable, theoretical entities in the twentieth century's most successful scientific theories has been too great to ignore. However, some may still hold the view that operational concepts should be used wherever possible, and may thus view the replacement of theoretical concepts with operational ones as a form of scientific progress[15].

It has been argued that the frequency theory *is* fundamentally operationalist in the sense that the probability is defined directly with reference to empirical observation. Indeed, Von Mises argued that the observation of large (but finite) empirical collectives provided a close enough approximation to his mathematical definition to be acceptable to operationalists. And there are other areas of natural science where empirically unobservable limits are used in definitions (most obviously, anything involving integral calculus). But this view was evidently not held by other leading philosophers with operationalist outlooks (Jeffrey's opinion on the frequency theory's inadequacy in light of operationalist requirements was also shared by other philosophers of probability of the era such as De Finetti, who we will also discuss further below)[16].

In contemporary writing, some anti-frequentist thinkers have argued that the difficulties with the empirical implementation and interpretation of the frequency theory are especially acute for very low-probability events[17]. In such cases, it has been argued, the observation of an extremely-unlikely event such as being dealt a club from a randomly shuffled deck of cards (with replacement) 1000 times in a row would lead to the immediate conclusion that there was something wrong with the experimental set-up, and hence such an event's probability could never be estimated in frequentist terms. But ultra-low probabilities are used regularly in physical science – for example, the probability that an alpha particle tunnels into the nucleus of an U-238 uranium atom is known to be around 10^{-38}. This probability has been estimated through the experimental observation of the behaviour of a huge number of alpha particles and uranium atoms. The frequency approach to probability appears perfectly suited to the definition of this empirically-observed ultra-low probability.

The Reference Class Problem

The second philosophical issue that arises with the frequency theory is often called the *reference class problem*. Recall above that Von Mises' definition referred to a collective as 'a sequence of uniform events or processes which differ by certain observable attributes'. What does it mean for two events to be uniform but different? Clearly, the events cannot be uniform in every respect, otherwise they would each result in the same outcome and the probability could only be 0 or 1 – so there must be some form of variation in the attributes of the 'uniform events' in order for non-trivial probabilities to be obtained. In Venn's definition, he described how the instances of a series would have a number of attributes in common (which define them as belonging to the series), and there would also be some attributes that are found in some instances and not in others (and these attributes arise in a 'certain definite proportion of the whole number of cases'). These parts of Von Mises' and Venn's definitions are essentially similar. So, in Von Mises' terminology, in order for an event to qualify for inclusion in a collective, each event must be uniform in the sense that it has attributes $B_1, B_2, \ldots, B_n$. And of these events that have those attributes, some portion will also have attribute A, and the remainder will not. The collective, and hence the probability of A, changes with the choice of 'uniform' attributes B. If we define the collective with reference to B_1 only, the probability of A is different to when we define the collective with reference to B_1 and B_2. The

---

15 For a fuller discussion of the philosophical doctrine of operationalism, see, for example, Machlup (1978), Chapters 6 and 7.

16 See Gilles (2000), p. 97-104 for a fuller discussion of operationalist concerns about the Venn limit.

17 See, for example, Jaynes (2003), p. 323.

reference class problem is concerned with the apparent arbitrariness of the choice of the uniform attributes that determine the collective.

This issue will of course be familiar to actuaries, demographers, insurance underwriters or indeed anyone who has thought about the identification of homogenous risk groups. One of the most influential (philosophical) pieces addressing this point was written by the Oxford philosopher A.J. Ayer in 1963[18]. He used an actuarial example to illustrate his point. He considered the probability that a man survives to age 80. This probability must be measured with reference to some class of lives. But which class? Which set of attributes do we assume are uniform across the set? The class could refer to any or all of characteristics such as sex, profession, smoking habits, drinking habits and present health and so on. The choice must involve a seemingly arbitrary decision about the granularity of the reference class. Intuitively, the class that *includes the most relevant information that is available* would seem the 'right' probability and Ayer's philosophical reflections led him to this conclusion – he argued for the use of 'the narrowest class in which the property occurs with an extrapolable frequency'[19].

Ayers was far from the first philosopher to raise the reference class problem. Nor was he the first philosopher to consider the rule of using the narrowest available reference class – it was mooted by Venn one hundred years earlier. Later we will discuss John Maynard Keynes' approach to probability. Suffice it to say for now, he did not subscribe to the frequency theory, and in his view, the difficulties it had with the reference class problem led him to conclude they 'appear to me insurmountable'[20]. He explicitly rejected the law of the narrowest reference class, highlighting that the logic implied a reference class of 1:

"If the process of narrowing the class were to be carried to its furthest point, we should generally be left with a class whose *only* member is the proposition in question, for we generally know something about it which is true of no other proposition."[21]

Ayers' caveat of 'extrapolable frequency' attempts to address Keynes' point, but the ambiguity of this term and the absence of an explicit definition by Ayer arguably merely re-states the problem of arbitrariness in a different form.

Ayer's approach to the reference class problem is, arguably, an *epistemic* one: he advocated using the narrowest reference class that is *known in the given situation*. That is, the reference class is *epistemically relative*. This relativity may be philosophically unappealing to objectivists – recall above how Popper regarded probability as a real property of a physical object. Such a perspective leaves little room for epistemic relativity. Some objectivists may therefore prefer to argue that there are some physical phenomena that are *irreducibly statistical*, even when given all information relating to the phenomena. This is sometimes referred to as *ontological indeterminism*[22]. Classical Newtonian physics and the Laplacian determinism that accompanied it is incompatible with the notion of ontological indeterminism. But quantum physics, which was probably the most successful physical scientific theory of the twentieth century, has irreducibly statistical descriptions of physical phenomena at the core of the theory. Quantum physics provides probabilistic descriptions for the behaviour of sub-atomic particles such as photons. When considered in large volumes, these descriptions can provide predictions that are consistent with classical physics. But quantum physics

---

[18] Ayer (1963)
[19] Ayer (1963), p. 202.
[20] Keynes (1921), p. 103.
[21] Keynes (1921), p.103.
[22] Ben-Haim (2014)

does not make deterministic statements about the future behaviour of individual particles, even when the current state of the physical system is fully specified. Dirac went so far as to say that, when considering the behaviour of an individual photon, 'it cannot be investigated and should be regarded as outside the domain of science'[23].

When a phenomenon is irreducibly statistical, its narrowest reference class will not be reducible to a single instance in the way Keynes suggested. Instead, in the above terminology, the physical system will have a finite number *n* of attributes, $B_1, \dots B_n$, and the relevant reference class will be of a size greater than 1 even when all of these attributes are known and specified (and it may indeed be of infinite size). So, from this perspective, in the case of recurring physical phenomena (such as the radioactive decay of atoms) there will be an *objectively homogenous* reference class of unlimited size[24]. The concept of objective homogeneity in the presence of irreducibly statistical phenomena fits the frequency definition extremely well, providing a form of solution to the reference class problem. However, there are (at least) two potential rebuttals of this argument: some philosophers may reject the idea that *anything* is irreducibly statistical (that is, ultimately every process is deterministic given sufficient knowledge of its state); second, if the frequency definition can only apply to irreducibly statistical physical phenomena such as radioactive decay, this seems highly restrictive and rules out many intended uses of the notion of probability.

Objective Probability and Singular Events

The reference class problem and the Venn limit are two major areas of philosophical concern with the frequency definition of probability. A third major philosophical topic that can arise with the frequency theory is the difficulty in applying it to singular, unique events. As discussed above, the frequency theory defines a probability for a collective, i.e. for a set of a very large (or infinite) number of events that share some specific common attributes. *The probability applies to the collective, not to the events that constitute it*. There are, however, many uses of probabilities that cannot be related to this idea of an infinite series of recurring, uniform events. For example, what is the probability of my horse winning next year's Grand National? What is the probability that a given scientific hypothesis is true? In Venn and Von Mises' views, quantitative probability is not applicable to such cases – there is no objective probability that can be assigned to these singular events, and that is essentially the end of the matter. As far as they were concerned, the probability does not meaningfully exist for these specific events. Von Mises used the mortality rate of an individual to emphasise the point:

"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 [rather than a collective of uniform lives] has no meaning at all for us."[25]

This need not necessarily offend traditional actuarial thought on mortality modelling, which is of course based on the estimation of probabilities that are applicable to reasonably homogenous groups rather than specific individuals. But it is more challenging to fit such a theory with, say, approaches to estimating the probabilities associated with specific extreme insurable events (say, a particularly strong hurricane next year in Florida).

The frequency theory's trouble with singular events can be related back to Keyne's perspective on the reference class problem. Recall, for a final time, Von Mises' definition of a collective as 'a

---

[23] Dirac (1958), p. 6. See Ben-Haim (2014) for further discussion.
[24] See Salmon (1984), especially Chapter 3, for an extensive philosophical discussion of these terms and ideas.
[25] Von Mises (1928), p. 11.

sequence of uniform events or processes which differ by certain observable attributes'. We saw that the essence of the reference class problem is that we need to decide which attributes are used to determine what defines the uniform events (for example, all lives are male and aged 50 at a particular date), and hence which attributes may be different (everything else), and there is no obvious objective rule for making this partition. As we move attributes from the 'different' category into the 'uniform' category, we may reach the limiting case where the collective of uniform events consists of one singular, unique non-repeatable event which has a probability of 0 or 1.

Below we will consider another theory of objective probability that aims to address the three philosophical concerns outlined above (the Venn limit; the reference class problem; and the applicability to singular events) whilst still fitting squarely within the objective approach to probability. Before we move on from the frequency theory however, it may be interesting to note what the philosophers who have advocated a frequency approach have thought about whether and how objective frequency-based probabilities can be found empirically. Venn, in particular, was an empiricist by philosophical outlook. In Venn's view, the series of events that have the characteristics postulated in his frequency theory would rarely in fact be found in nature and would tend to only arise in artificial games of chance. He was therefore sceptical about the value and reliability of empirical statistical analysis. He argued this was because nature's conditions were rarely, if ever, static. Rather, they tend to be subject to what Venn termed 'an irregular secular variation'. As a result, there was 'really nothing which we can with propriety call an objective probability'[26]. To Venn, an empirical time series would almost always be ultimately non-stationary:

"That uniformity which is found in the long run, and which presents so great a contrast to the individual disorder, though durable is not everlasting. Keep on watching it long enough, and it will be found almost invariably to fluctuate, and in time may prove as utterly irreducible to rule, and therefore as incapable of prediction, as the individual cases themselves."[27]

Today, it is widely accepted that stable probabilistic universal laws exist in the natural sciences (for example, in the field of radioactivity mentioned in examples above). But the form of non-stationarity described by Venn is a perennial problem in the social sciences (changes in life expectancy; behaviour of inflation rates, etc.) The theme will therefore recur in the discussion of modelling in the social sciences later in Chapters 3 and 4, and indeed in the discussions of actuarial science and modelling in Chapters 5 and 6.

From Frequency to Propensity

The *propensity theory* of objective probability can be viewed as a modification of the frequency theory that is primarily designed to permit the existence of probabilities for singular events, whilst remaining firmly within the philosophical framework of objective probability. It has numerous versions, perhaps the most notable of which was developed by Sir Karl Popper, who was one of the most influential philosophers of science of the 20th century[28].

The essential idea of the propensity theory is that it goes a step further than the frequency theory by specifying that the events within a collective are generated by the same set of repeatable conditions:

"The frequency interpretation always takes probability as relative to a sequence which is assumed as given....but with our modification, the sequence in its turn is defined by its set of generating

---

[26] Venn (1888), p. 91.
[27] Venn (1888), p. 14.
[28] Popper (1959).

conditions; and in such a way that probability may now be said to be a property of the generating conditions. But this makes a very great difference, especially to the probability of a singular event. For now we can say that the singular event possess a probability owing to the fact that is an event produced in accordance with the generating conditions, rather than owing to the fact that it is a member of the sequence."[29]

Popper argued that the propensity theory's idea of a set of underlying generating conditions allowed the event to be uncoupled from its sequence or collective, and therefore allowed the probability property to be associated with a singular event that belonged to the collective, rather than only to the collective itself. This detachment from the collective also removes the reliance on the Venn limit in the definition of the probability – the probability applies to a given event irrespective of how many times the conditions that generate the event are repeated, and so there is no necessity to consider repetitions that approach infinity in size. It is less clear how the propensity theory alters the frequency theory's difficulties with the reference class problem: any arbitrariness in the specification of attributes that defines uniformity of events in the frequency collective would seem to be equivalent to an arbitrariness in the specification of the generating conditions in the propensity theory.

It should be noted that, whilst Popper believed that the propensity theory would permit probabilities to be attached to singular events rather than only collectives, he did not believe that such a probability could be applied to *any* singular event. In particular, Popper argued that where 'objective conditions of the event are ill-defined and irreproducible', no objective probability can apply. As noted above, Popper required the generating conditions to be well-defined and *repeatable* in order for probability to apply. The notion of a *repeatable* singular event is not a straightforward concept and it does not have an obvious applicability to specific events such as my fishing trip this evening. So, from this perspective, it seems that the propensity theory only represents a very partial solution to the frequency theory's difficulty with singular events. Like Venn, Popper's position was that meaningful quantitative probabilities simply could not be associated with many forms of uncertain phenomena.

The remainder of Chapter 1 will discuss broader definitions of probability that attempt to encompass the broader set of circumstances in which we find probabilistic ideas arising.

1.2 Theories of subjective probability

In the previous section, probability was taken to be a characteristic of the physical world that was capable, in principle, of objective measurement. An objective probability may be unknown, or subject to measurement challenge, but it is not merely a perception of an individual: it *exists* as a property of a thing and this property is assumed to be independent of human knowledge. The frequency theory says probability is merely the frequency with which a certain attribute arises within a very large set of events that share some other defined attributes in common. When objective probability is considered from the propensity perspective, probability depends on the conditions associated with the event in question and these conditions are well-defined and repeatable features of the physical world. That is how an objectivist defines probability. It is a fairly intuitive perspective. But this perspective restricts the application of probability to mass phenomena and repeatable events. An objectivist may concede without contradiction (as Venn did) that probabilities defined in this way can rarely be found 'in the real world'.

[29] Popper (1959), p. 34.

In common language (and indeed in business, science and beyond), the term 'probability' is used much more widely than this. The objectivist definition places much of this usage outside of the mathematical theory of probability. An objectivist would argue that such uses of the term are generally not capable of mathematical meaning. However, other probability theorists have developed a much broader and more flexible approach to the meaning of probability that can encompass a wider set of applications (indeed, an unlimited field of applications) whilst still being mathematically well-behaved. Well-behaved here is generally taken to mean being consistent with the commonly accepted mathematical axioms of probability such as those developed by Kolmogorov[30]. That is, in this setting, and as alluded to in the quotation of Jeffreys in Chapter 1.1, probability is defined indirectly by its axiomatic relations, rather than explicitly.

Thus, if something behaves in a way that is consistent with the Kolmogorov axioms, that may be good enough for a mathematician or probabilist to allow it to be called a probability. But the philosopher may wish to find another explicit interpretation for the meaning of probability that is contrary to the objectivist perspective. The standard philosophical alternative to the objectivist approach is called the *subjectivist* approach. It takes a completely different starting point for what is meant by probability. Here, probability is interpreted as an individual's *degree of belief* in a given event occurring or being true. It is therefore always conditional on the knowledge and information available to the individual at a given point in time. In this setting, probability has an *epistemological* interpretation. It is a measure of the incompleteness of an individual's knowledge about the determinants of an event. It can be related to Laplace's famous demon of universal determinism - that is, the view that, with perfect knowledge of all relevant conditions and attributes, it can be known with certainty whether an event will or will not occur. But it is not necessary to subscribe to universal determinism in order to find a use for epistemic probability.

Philosophers and historians have a range of views on when this epistemological interpretation of probability first emerged. Some, such as Sir Harold Jefrreys, have argued it has been a part of probabilistic thinking since before Bernoulli's Law of Large Numbers was published in 1713. It is widely regarded as being implicit in Bayes' paper of 1763. Historians such as Lorraine Daston have argued it emerged as a distinct, rather than implicit, part of probability in 1837 in the work of Poisson[31].

Within this epistemological approach to probability there are two philosophically distinct perspectives. One is to define probability as the degree of *reasonable* or *rational* belief in the truth of a statement. In this case, the probability is uniquely defined by the rational and logical processing of the specified information that it is conditional on – the probability is not a subjective quantity that depends on the judgment of the individual. This is called *logical* probability. It is an attempt to enumerate inductive logic. In the other approach to epistemic probability, *subjective* theories of probability relax the rationality stipulation. Instead, these theories' starting point is that individuals can come to their own subjective assessment of such a probability. And, rather than developing logical rules for how individuals make such assessments, subjective theories merely attempt to infer what these subjective probabilities are from the actions of the individual.

The logical approach pre-dates the subjective approach, and some form of it has been in use since Bayes and Laplace independently developed approaches to statistical inference between 1763 and 1810. The key mathematical developments in the subjective theory of probability were developed independently and contemporaneously by Frank Ramsay and Bruno De Finetti in the 1920s and

[30] Kolmogorov (1950)
[31] Daston (1988)

1930s[32]. As we shall see below, Ramsay and De Finetti share the title of a fundamental theorem in subjective probability. Despite this, their philosophical outlooks differed. Ramsay viewed the frequency theory of objective probability as a perfectly reasonable approach to probabilistic reasoning in some circumstances. De Finetti's philosophy wholly rejected the objectivist's philosophical perspective in which probability was a physical property of something, arguing that such a view implied a 'mysterious pseudo-world'[33]. From a historical perspective, these developments in subjective probability in the 1920s and 1930s may be viewed as a reaction to the then-prevailing view that the limitations of the logical approach increasingly seemed insurmountable. We will discuss these limitations of the logical approach, and contemporary advances in addressing them, in Chapter 1.3. First, this section introduces the subjective theory of probability.

Discovering Consistent Implied Subjective Probabilities: Coherent Betting Odds

Individuals' subjective degrees of belief in uncertain future events would seem to be an inherently psychological phenomena – how can such a thing be observed and quantified? Perhaps the most obvious response is to ask the subject what they believe. But such an approach would be full of ambiguities – what is to motivate the subject to tell the truth? What if the subject does not fully understand what their own motives would be in the hypothetical circumstances in question? Self-avowal is considered of dubious reliability not just by subjective probabilists but by economists, historians and others. So, to the subjectivist, probability is determined by *observing decision-making actions* under uncertainty. Ramsay and De Finetti each independently proposed that subjective probabilities should be measured by observing individuals' (possibly hypothetical) decisions or actions in a particular way: *by observing the minimum odds that an individual required in order to induce them to bet on a given event*[34]. So, for example, if John required a minimum pay-out (including the return of his initial bet) of £3 to be induced to bet £1 now on the event that it will rain tomorrow at noon, his (current) subjective probability of it raining tomorrow would be 1/3.

The use of betting odds to imply subjective probabilities was not a completely new idea when it was advocated by Ramsay and De Finetti in the 1920s and '30s. Earlier works on the philosophy of probability such as Venn's *Logic of Chance* had discussed the idea[35]. But Ramsay and De Finetti both identified an important property of these subjective probabilities that elevated subjective probability to a level of mathematical respectability. This property became known as the *Ramsay-De Finetti theorem*: if the betting odds are specified in a way that avoids it being possible for the individual to construct a combination of bets that will generate a certain profit or loss (what an

[32] The key papers of Ramsay and De Finetti, along with other important papers on the topic of subjective probability, can be found in Kyburg and Smokler (1980). A comprehensive presentation of De Finetti's conception of probability theory and statistical inference can be found in De Finetti (1975).

[33] Hacking (1965), p. 211.

[34] Technically, the stakes of the bet must be assumed to be small relative to the overall wealth of the bettors so that complications arising from individual risk-aversion and utility preferences can be avoided. Ramsay, in particular, developed an approach that could allow for the individual's utility function when inferring the subjective probabilities from their betting requirements. This foreshadows the linkage between subjective probability and asset pricing. Historically, the interrelationship between these two disciplines appears to have been significant. For example, it was Leonard Savage, the leading subjectivist probabilist, who first introduced Paul Samuelson, the economist, to the stochastic modelling work of Bachelier that arguably provided the genesis of the Black-Scholes-Merton equation. Savage played an important role in popularising the subjective approach to probability during the third quarter of the 20th century. See Savage (1972).

[35] Venn (1888), p. 144-6.

economist would call an arbitrage), then the subjective probabilities will satisfy all the standard mathematical axioms of probability. In particular:

- The subjective probabilities of a set of mutually exclusive and exhaustive events will sum to 1 (the Addition Law). So, in the above example, John's subjective probability that it is not raining tomorrow at noon must be 2/3 (which would be implied by requiring a pay-out of £1.50 in the event of it not raining at noon tomorrow).
- The subjective probability of two events, A and B, jointly occurring is equal to the probability of A occurring multiplied by the conditional probability of B occurring given A occurs (the Multiplication Law). Note that this requires the subject to quote betting quotients on *joint* bets and *conditional* bets (in the conditional case, the bettor would have his stake returned if the conditioning event did not occur). So, to extend the above example, suppose John also holds a subjective probability for the joint occurrence of the event that it rains tomorrow at noon *and* that Red Rum wins that afternoon's horse race of 1/5 (and so the joint bet pay-out is £5). Then his subjective probability that Red Rum wins *given* it will rain tomorrow at noon must be 3/5. And so the pay-out on a £1 conditional bet must be £1.67 (or exactly five-thirds of £1) in order for the bets to be arbitrage-free[36].

Equivalently, if the subjective probabilities for a set of events satisfy the axioms of probability, then the bets will be arbitrage-free. That is, Kolmogorov's axioms (or similar) are necessary and sufficient conditions for the arbitrage-free property. This arbitrage-free property is referred to by subjectivists as *coherence.*

There are, perhaps inevitably, some conceptual difficulties associated with the framework of discovering subjective probabilities by observation of betting odds. What if there is a possibility that the event in question may never be conclusively known to be true or false? For example, does it make sense for an individual to bet on the probability of a scientific theory being true, when it may take many hundreds of years (or, indeed, forever) to establish if it is true or not? After all, Democritus first proposed a theory of molecular reality around 400 BC. Such a theory (albeit in somewhat more sophisticated form), was not widely accepted by the scientific community until the early twentieth century[37]. De Finetti argued that a probability could only be associated with a statement or event that was *well-determined* or *unequivocally individuated*, by which he meant that 'a possible bet based upon it can be decided without question'[38]. This restricts the domain of applicability of subjective probability, thereby excluding the use of probabilities from some forms of statements that are objectively (or physically) true or false.

The betting argument also assumes there is always a price that exists such that an individual will be willing to participate in a bet on such an event (unless it is known to be impossible). That is, the individual cannot merely say, for example, they have no knowledge or understanding of the event and hence refuse to consider the bet[39]. This is an important assumption – as we shall see below, the determination of probabilities in conditions of total ignorance has been a philosophically

[36] In the event it does not rain (probability of 2/3), the bettor's stake of £1 is returned. In the event it does rain and the horse wins (probability of 1/5), the bettor receives £1.67. In the event it does rain the horse loses (probability of 7/15), the bettor receives zero. Note the average pay-out (including return of stake) is £1, i.e. the size of the initial bet.

[37] The interested reader is referred to Nye (1972) for an excellent account of the history of scientific theories of molecular reality.

[38] De Finetti (1975), Chapter 2.3.

[39] See De Finetti (1975), Chapter 3.3.3.

contentious area that has sometimes been identified as the Achilles heel of some definitions of probability (logical probability, in particular).

Finally, another complication, as briefly noted above, is that we must untangle utility and risk-aversion from the subjective probability implied by the bet size. Those schooled in financial derivative pricing theory will be familiar with the idea that prices alone can only imply risk-neutral, rather than actual ('real-world') probabilities. The same logic applies here. De Finetti argued that the risk-aversion complication could be avoided by assuming the bet sizes were small relative to the individual's overall wealth[40].

So, to summarise. We noted above that subjective probability theory differs from logical probability theory in the sense that the subjective theory does not attempt to establish a rational method for determining probabilities as degrees of belief. The subjective theory places no limits on how the probability of any given event is assigned by an individual. But some form of rationality restriction is necessary if the related probabilities are to be collectively 'well-behaved' (that is, to comply with probability's standard mathematical axioms). This restriction is the subjectivist's concept of coherence, which is essentially the same as the economist's concept of no-arbitrage, and it is encapsulated by the Ramsay-De Finetti theorem. The theorem is regarded as important because it means that, irrespective of philosophical outlook or the epistemological interpretation that is attached to subjective probabilities, the theorem makes the subjective perspective an unambiguously valid way of interpreting the mathematic calculus of probability. Nonetheless, there are some noted conceptual difficulties with the universal use of betting odds in the discovery of subjective probabilities.

Exchangeability

As noted above, the conditionality of subjective probabilities on the available evidence implies that, to the subjectivist, there is no 'true' or singular underlying probability distribution: each new piece of evidence gives rise to an updated probability distribution. The notion of a statistically independent set of identically-distributed trials therefore does not arise in subjectivist probability and statistics: the estimate of the underlying probability distribution is continually updated as the evidence from each new observation emerges, and so each new observation has a different subjective probability distribution. But assumptions of independence and identically distributed samples play a critical role in how objectivists are able to use sample data to make statistical inferences – don't subjectivists need to make some similar types of assumptions about the nature of the sample data when updating probability distributions?

The answer is yes, *some* form of assumption is necessary, but there is no logical constraint on what form of assumption is necessary (or appropriate). The simplest case is where the data can be considered as *exchangeable*. The concept of exchangeability can be loosely thought of as the subjectivist's equivalent of the objectivist statistician's assumption of independent, identically-distributed trials. The exchangeability condition is met if the subjective probability conditional on observing r events in a sample of n observations is the same for all permutations of the order in which the r events arose (i.e. the sequence or path of the observations is irrelevant). De Finetti described this distinction between independence and exchangeability using an example of sampling marbles from an urn:

"Extractions with return of the marble to the urn having a *known* composition (e.g. 7 white marbles and 3 black ones) are stochastically independent; if, on the other hand, the composition is *not*

---

[40] See De Finetti (1975), Chapter 3.2.

known…it is clear that each piece of information on the results of a new extraction increases the probability assigned to the composition of the colour of the marble extracted. Independence thus does not exist: the correct translation of the nonsensical 'constant but unknown probability' is given by the notion of exchangeability. In these cases, later extractions are (not independent but) exchangeable in the sense that the probability does not vary for permutations e.g. every series of 9 extractions (e.g. of 5 white and 4 black) has the same probability."[41]

Like the objectivist's assumption of independent and identically-distributed, the subjectivist assumption of exchangeability may or may not be applicable to a given set of data. Moreover, it is arguably less clear when the exchangeability assumption is not appropriate. The contemporary philosopher of science and probability Donald Gillies (who has definite objectivist leanings) has argued that the only way to test if the exchangeability assumption is reasonable for a given set of data is to test if the data is objectively independent. He argues that the concept of exchangeability is consequently 'parasitic on objective independence and therefore redundant'[42]. Nonetheless, if probability is defined subjectively rather than according to an objectivist theory, exchangeability is a fundamentally important concept.

Updating subjective probabilities

When armed with the exchangeability assumption, the updating of subjective probabilities to reflect new observational data is fairly straightforward. The updating process may be a lot less straightforward when new information is not of the form of well-defined events that meet the exchangeability criteria. All we can really say about this general situation is that it depends on the specifics of the case. Moreover, in the case of exchangeable data, the updating process makes no distinction between the two epistemological probability branches of subjective and logical probability – the subjectivists' *coherence* constraint determines a unique way of updating probabilities in the presence of new exchangeable data, and this updating method is also the only one, in these conditions, consistent with the logical approach.

The basic updating mechanism goes as far back as Bayes' paper of 1764 in which Bayes' Theorem was first set out. In mathematical terms, the standard rendering of Bayes' Theorem is:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

Despite the controversy surrounding the use of Bayes' Theorem, the formula itself is quite unremarkable and is easily derived from Kolmogorov's axioms (which is one of the few things that objectivist and subjectivists can mutually accept). When the three probabilities on the right-hand side of the equation have known values, there is no logical ambiguity about the applicability of the formula. The objectivists' objection is focused on how Bayes' Theorem is applied to the problem of statistical inference, and in particular whether all three of these probabilities can be meaningfully quantified in that context. For now, let us merely note this is a topic of contention that we will return to in Chapter 1.3.

The iterative process of updating a subjective probability for each new piece of exchangeable observational data is sometimes referred to as *Bayesian conditioning* (and in this context Bayes' Theorem may sometimes be referred to as the *Principle of Inverse Probability*, a term introduced by Laplace). We start with a subjective probability of some event A, P(A), conditional on all currently

---

[41] Kyburg and Smokler (1980), p. 210.
[42] Gillies (2000), p. 77.

available information. We obtain new information, B. Bayes' Theorem is then used to produce a posterior probability, P(A|B).

The posterior distribution given *n* pieces of data becomes the prior distribution for the (*n+1*)th piece of data. That is:

$$P(A|e_{n+1}) = \frac{P(A|e_n)P(e_{n+1}|A)}{P(e_{n+1})}$$

This process can be iterated again with the next data observation to produce a new posterior distribution. And so on *ad infinitum*. Importantly, as a greater volume of data is introduced, the posterior distribution converges on a result which is independent of the initial starting assumption for the prior distribution (assuming the evidence meets the exchangeability criteria described above by De Finetti; and providing the starting prior probability is not zero or one).

Objectivists such as Donald Gillies have argued that, when applied to a time series of data, exchangeability and Bayesian conditioning may attach too much weight to older data. That is, the posterior probability may not change quickly enough in light of important new evidence. Gillies offers an example[43]. If the sun was observed to rise for one million days in a row and then did not rise today, the posterior probability produced for the sun rising tomorrow would be counter-intuitively high when we assume that the million and one days of data is exchangeable. Whilst this criticism is doubtless reasonable, it must be noted that objective probability has a very similar failing: if we assume the million and one days of data are independent and identically-distributed, we would again infer the probability of the sun rising tomorrow is bizarrely high. Neither the objective nor subjective approaches offer an obvious solution or strategy for handling this (rather contrived) example.

Whilst the posterior distribution produced by a very large volume of data does not strongly depend on the assumed size of the initial prior probability, it will be an important assumption in the context of smaller data sets. As we have seen above, subjective probability theory has little to say about how these initial or prior probabilities are generated, other than that they can be observed from hypothetical betting actions and will obey the basic axioms of probability when the betting is arbitrage-free (coherent). We will see in the next section how the logical approach to probability attempts to say more about the specification of these probabilities, especially in the condition of a total lack of information.

Intersubjective probabilities

So far, the discussion of subjective probabilities has considered the probabilities of an *individual*. What if *multiple* individuals have subjective probabilities for the same event. Interestingly, it was some half a century after the breakthrough work of Ramsay and De Finetti that someone (the Australian actuary, J.M. Ryder) pointed out that if different individuals used different subjective probabilities (as implied by their betting quotients), then this implied a third-party could arbitrage between these bets by selling the short-odds bet and buying the long-odds bet[44]. We saw how the arbitrage-free condition was fundamental to demonstrating that subjective probabilities could be made to satisfy the basic mathematical axioms of probability. Ryder pointed out that the arbitrage-free condition also implied that the notion of subjective probabilities as completely individualistic

---

[43] Gillies (2000), p. 73.

[44] Ryder (1981). It may be of particular interest to the actuarial audience to note that the philosopher responsible for this breakthrough and the foundation of intersubjective probability was a Fellow of the Institute of Actuaries and a former Government Actuary of Australia.

belief may be inadequate. Ryder held a strongly objectivist outlook, and he argued that this form of arbitrage-free condition meant the entire notion of determining subjective probabilities with regard to an individual's betting quotients was not viable. However, some philosophers have suggested an alternative interpretation of Ryder's arbitrage-free constraint: that it gives rise to *intersubjective* probabilities, where the subjective probabilities are essentially consensus group probabilities rather than entirely individual ones.

The intersubjective approach gives subjective probability a social as well as psychological element, and the idea arguably further strengthens the link between subjective probability and the economics of risk and asset pricing – financial market prices are perhaps the most natural example of where we might attempt to infer intersubjective probabilities. Indeed, one argument in favour of the use of (inter)subjective probabilities implied by market prices is that the events that financial market prices are contingent on clearly do not have objective probabilities (in the sense that these events are not mass phenomena produced by repeatable events or recurring generating conditions). To the extent that probabilities are used at all in the consideration of such claims, these probabilities can only be (inter)subjective.

Probability and Statistics: Objectivist and Subjectivist Perspectives

Let us now take stock of where we have got to thus far by comparing and contrasting the fundamentally different outlooks of objective and subjective approaches to probability and statistics.

Subjective probabilities, as degrees of belief, are fundamentally conditional on what is known to the given subject at a given point in time, i.e. on the evidence available to support belief in the truth of the statement in question. To the objectivist, on the other hand, a probability exists as a real, objective, 'physical' property of something. The probability does not vary as a function of an individual's knowledge. Where the magnitude of the probability is unknown, sample observations can be used, under suitable assumptions, to infer estimates of this objectively true, constant but unknown quantity. This is the objectivists' basic perspective on statistical inference.

In objectivist statistics, the canonical technique for estimation of the parameters of a probability distribution from some (independent and identically distributed) sample data is maximum likelihood. This approach aims to identify the population probability parameters that attach the highest probability to the occurrence of the observed sample distribution. Its development is most strongly associated with the work of Sir Ronald Fisher in the early twentieth century, who demonstrated the inferential properties of maximum likelihood[45]. However, as a concept, it arguably goes as far back as Daniel Bernoulli's writings of 1777[46]. To take a simple example, suppose we know that a population has a normal distribution with a known variance but unknown mean. A sample of size n is observed from this population. It can be easily shown that the maximum likelihood estimate of the population mean is the arithmetic mean of the sample (assuming the sample is of independent and identically distributed trials). Note that the maximum likelihood method, and similar objective approaches to statistics, will assume a form of prior knowledge: that is, that the population from which the sample is taken is known to belong to some particular form or family of probability distributions or stochastic processes. It may be convenient to make mathematically tractable distributional assumptions, but of course, such assumptions may or may not be reasonable

---

[45] Over the course of his career Fisher produced an enormous volume of paper and books on statistical theory and practice, but for a good example of his foundational thinking on the method of maximum likelihood, see Fisher (1922).

[46] See Hacking (1965), p. 63-4.

in specific circumstances. Note also that the required assumption is not merely that the population probability distribution is 'probably' of a given form, but that it is *known* to be.

The other key plank of objectivist statistics is hypothesis testing. The general strategy here is to conjecture that a given state of the world is true (for example, a null hypothesis that the mean of a population probability distribution is zero), and then determine if the sample data evidence is strong enough to reject this assertion with a high degree of confidence. This is known as significance testing or hypothesis testing. Fisher was again highly influential in the development of the theory and practice of hypothesis testing, but he had notable rivals in others such as Neyman and Pearson[47]. Like any form of statistical inference, hypothesis testing is fallible. Two forms of error can arise: the error of rejecting a true hypothesis (Type I error); and the error of not rejecting a false hypothesis (Type II error; the 'power' of the test is defined as the probability of rejecting the hypothesis when it is false). The probability of a Type I error is controlled by the significance level of the test. The probability of the Type II error will generally decrease as the significance level (and the probability of Type I error) increases and vice versa (all else being equal). Under suitable conditions, the probability of the Type II will reduce asymptotically to zero as the sample size increases.

From a subjectivist perspective, however, objective 'physical' probabilities do not exist. In this setting all probability essentially becomes statistics in the sense that a constant, unknown probability has no meaning: probability is defined by and is explicitly conditional on the (limited) available information (such as some sample data). Probability is therefore not a constant unknown but an epistemic property that is continually changing as new information arises. In subjectivist (or Bayesian, as it is more commonly referred to) statistics, the population probability distribution parameter(s) itself has a probability distribution which is updated to reflect new observations. A subjectivist approach to the simple statistics example given above of estimating the mean of a population normal distribution with known variance would assume a starting (prior) probability distribution for the mean of the distribution, and then update this distribution following observations to obtain a posterior distribution. The subjectivist's estimate of the population mean would usually then be defined as the median or mean of the posterior distribution. In this example, given our earlier assumption that the data independent and identically distributed, it has the subjectivist property of exchangeability, and it can be easily shown that the posterior distribution will converge on the population mean as the sample size increases.

The objectivist and subjectivist approaches to the statistical task of interval estimation also contrast. An objectivist approach will employ a confidence interval, which states that the true, constant but unknown population parameter is within a specific range with a given level of probability. The theory of confidence intervals was formally developed by Neyman and can be considered as the natural counterpart of hypothesis testing[48]. The 'given level of probability' referred to here has a frequentist interpretation in the sense that if the sampling procedure is repeated, with a new confidence interval produced each time, the true (constant) population parameter will appear within the calculated interval with a frequency that will tend towards the stated confidence level as the number of samples tends to infinity. As the size of the data sample increases, the confidence interval will narrow (under suitable conditions). A subjectivist would naturally use the posterior probability distribution to make interval estimates for the parameter (and in Bayesian statistics these are called

[47] Neyman and Pearson (1933)
[48] Neyman (1937)

credibility intervals). Again, as the data sample size increases, the posterior distribution, and credibility interval, will narrow.[49]

From this brief overview, it is apparent that objectivist and subjectivist approaches to statistical inference differ significantly in philosophical outlook, in the logic of the inferential process and in the jargon employed. The good news, however, is that for most standard types of problem, they tend to ultimately converge on the same or very similar answers!

Next, we consider the third and final philosophical branch of probability definitions: the logical theory.

1.3 Induction and the logical approach to probability

The logical theory of probability, like the subjective theory, starts with the epistemological approach of defining probability as a measure of the degree of belief in the truth of a statement. The logical theory, like the subjective theory, therefore views probability as being fundamentally conditional on available knowledge. Logical theories of probability can be viewed as a special case of the subjective approach where the degree of belief is defined in a particular way: in subjective probability, the probability is inferred from an individual's actions without making assumptions about the individual's processes for arriving at the probability other than that they will avoid the possibility of being arbitraged (coherence). But in the logical approach a stronger assumption is made - the individual's degree of belief is assumed to be that which is *reasonable* or *rational* to hold, given the information available. So, the logical approach aims to be more objective than subjective in the sense that it assumes that (in at least some cases) there is a unique and objectively-determined probability that is the result of a rational assessment of the evidence. Different individuals will therefore *not* hold different logical probabilities, given the same information, whereas in subjective probability theory the probability is only constrained by the assumption of coherence (intersubjective probability theory also implies individuals will not hold different subjective probabilities, which makes this distinction less clear-cut).

Given the above, the logical probability approach naturally gives rise to the same prior / evidence / posterior updating structure that is present in subjective theories of probability. As was noted above, the two approaches do not differ in how to update prior beliefs when new, well-behaved (i.e. exchangeable) data is received. The Bayesian conditioning process described above for subjective probabilities similarly applies to logical probabilities. The starkest difference between the two approaches can be found in how *prior* probability distributions are specified. We will discuss this topic in some depth later in this section.

On Induction

Advocates of the logical approach such as John Maynard Keynes, Rudolf Carnap and Harold Jeffreys viewed logical probability as a form of inductive logic that could sit alongside traditional, deductive logic as part of a theory of knowledge. Keynes viewed the frequentist definition of probability as a small sub-set of logical probability: "The cases in which we can determine the logical value of a conclusion entirely on grounds of statistical frequency would seem to be extremely small in number."[50]. Carnap drew a sharp distinction between logical and frequentist probabilities, referring to the former as 'probability$_1$' and the latter as 'probability$_2$'[51]. Meanwhile, as will be discussed further below, Jeffreys rejected the frequentist definition of probability altogether.

---

[49] See Howson and Urbach (1993) for a fuller discussion of the Bayesian approach to statistical inference.
[50] Keynes (1921), p. 99.
[51] Carnap (1950), Chapter II.

To gain some appreciation of the logical perspective on probability, a brief diversion into the topic of induction and inductive knowledge will be helpful. In epistemology, the process of inferring knowledge from evidence can be considered to arise in two basic ways: by deduction and by induction. A deductive inference takes premises that are known (or assumed to be known by hypothesis or postulate) and applies laws of logic to arrive at a conclusion that is certain, given the truth of the premises (the conclusion is said to be logically entailed by the premises).

Deductive logic classically takes the form of a *syllogism*. For example: 1 All actuaries know how to linearly extrapolate; 2 Craig is an actuary; Conclusion, Craig knows how to linearly extrapolate. Whilst seemingly logically straightforward, the syllogism has nonetheless generated its own literature of philosophical criticism. For example, Venn pointed out that if we knew that Craig was an actuary when we stated the syllogism, there really was no need to state both premises. Whereas, if we did not already know Craig was an actuary when we established the first premise, then how could we be so confident that it was indeed true?[52]

In an inductive inference, the conclusion is not logically entailed by the premises, but the premises (if true) may nonetheless provide positive (but inconclusive) supporting evidence for the conclusion. They might even be said to make the conclusion *probable* (or at least *more probable* than if the premises were false). Venn's criticism highlights the path from a syllogism to an inductive inference. For example: 1 I have observed some actuaries (but not Craig) and all of those observed actuaries can linearly extrapolate; 2 It is reasonable to expect that all actuaries (including the unobserved ones) can linearly extrapolate; 3 Craig is an actuary; Conclusion, It is reasonable to infer Craig probably knows how to linearly extrapolate.

These two forms of argument are clearly fundamentally different. The deductive syllogism takes a universal generalisation as known, and this then enables definite statements to be made about specific cases. The inductive argument takes some specific cases (such as a finite set of observations) as known and then *infers* a universal generalisation which then allows statements to be made about other specific cases not in the set of observations. Inductive inference inevitably results in a form of knowledge that is fallible – the truth of the conclusion is not logically certain, but is contingent on the reliability of the universal generalisation that has been hypothesised from some limited initial information such as a finite set of prior observations. Epistemologists generally regard inductive inferences as a legitimate form of justified knowledge, with the general caveat that the greater and more representative the set of prior observations, the greater the justification for that knowledge. Logical probability is essentially concerned with producing a mathematical apparatus that can provide a quantification of this degree of justification.

The inescapable fallibility of inductive inference was brought into sharp relief by the Edinburgh philosopher David Hume in the mid-18th century[53]. Hume's arguments highlighted the critical point that an inductive inference had to rely on more than just past observational data. That is, it also required some form of *non-empirical assumption* that gives a reason why the past (or observed) data is relevant for predicting behaviour in the future (or in the unobserved).

In general terms, this assumption can be described as assuming some form of uniformity of nature across time and space – such an assumption provides the explicit logical basis for the use of prior observations in inferring generalisations that are applied to similar non-observed events. The need for this assumption leads to a difficulty. How can this uniformity assumption be justified? The only

---

[52] Venn (1889), p. 373.
[53] Hume (1740), Hume (1748).

basis for its justification would appear to be that it has been found to usually work in other inferences. But this is merely an inductive inference – a finite set of observations of inductive inference working well in the past is used to infer that inductive inference will generally work in the future - and so leads to a circularity. The satisfactory justification of inductive inference surely cannot be grounded on an inductive inference. In philosophy, this lack of a rigorous logical justification for induction is known as the problem of induction. As Bertrand Russell succinctly concluded in his famous essay, *On Induction*, "Thus all knowledge which, on the basis of experience tells us something about what is not experienced, is based upon a belief which experience can neither confirm nor confute"[54].

And yet, in real life, we use inductive inference all the time. Deductive knowledge is limited to the domains of maths and logic. All of our other knowledge – including scientific knowledge - relies on limited experience and some form of inductive inference from that experience. But empirical observations, in and of themselves, have no informational content irrespective of their number unless accompanied by a non-empirical assumption of uniformity that is difficult or impossible to empirically justify.

Statistical inference can be viewed as a special, quantitative case of inductive inference. Statistical methods will make the assumption of a form of uniformity of nature explicit – by assuming that the sample of past observations are independent and identically distributed from a known distribution which also applies to unobserved population (if an objectivist). Without such assumptions, the inference from data has no logical basis.

Since the time of Hume, philosophers have invested much energy in attempting to solve the problem of induction – that is, to produce a philosophical justification for why empirical observation is informative without recourse to a non-empirical and philosophically unjustified assumption[55]. Much of this work hinged on linguistic contortions – for example, justified, probable belief could simply be defined as the fallible knowledge resulting from inductive inference under the assumption of a uniformity of nature. Some philosophers also argued that the problem of induction was a natural and inevitable feature of induction – inductive inference inevitably could not be proven to be correct, as that would make it deduction rather than induction. Another argument for the philosophical justification of induction was that it would work better than any other method when the assumption of a uniformity of nature held true, and that no inferential method could succeed when the assumption did not hold. None of these approaches 'solved' the problem as stated and today the problem of induction is generally regarded by philosophers as insoluble.

Despite a lack of philosophical justification for induction, inductive inference is still generally regarded as capable of producing knowledge – a particular kind of knowledge that is fallible, tenuous and subject to revision, which may or may not be true, but that nonetheless is to some extent justified and reasonable. It may be called probable knowledge (and this would seem to be a very important category of human knowledge, as it pertains to virtually all scientific knowledge. The topic of the scientific method as a formal model of the development of probable knowledge is discussed further in Chapter 2). This notion of probable knowledge gives rise to the use of probability as a measure of its strength - the degree of rational belief in the truth of a statement. This is the starting point for the logical theory of probability.

---

[54] Russell (1912)

[55] See, for example, Swinburne (1974).

A Swan-based Example of the Problem of Induction

The essence of logical probability and its relationship to inductive inference can be illustrated with a very simple example. In time-honoured tradition, let us consider the colour of swans in our inductive illustration[56]. We begin with a hypothesis that conforms to our observational experience to date: 'All swans are white'. Let's call this hypothesis *A*. Suppose there is a swan in the garden and we are about to look at it in such a way as to correctly determine its colour. We will denote the colour of this swan by a binary term *B*, where *B* = true if the swan in the garden is white, and *B* = false if the swan in the garden is any other colour (black, blue, pink, yellow, etc.). Before we set our sights on the swan in the garden, we attach some probability to the truth of hypothesis *A*, which we will denote as P(A). Self-evidently, if *B* is false, then *A* is false. That is:

$$P(A|B = false) \ = 0$$

Inductive inference is concerned with how we update P(A) in the event that B = true. That is, what is the relation between P(A) and P(A|B = true)?

A simple application of Bayes' Theorem tells us:

$$P(A|B = true) \ = \frac{P(A)P(B = true\ |A)}{P(B)}$$

Clearly, P (B = true | A) = 1. That is, if it is true all swans are white, we can be certain that the swan in the garden is white. This equation therefore simplifies to:

$$P(A|B = true) \ = \frac{P(A)}{P(B)}$$

Noting that $P(B) \leq 1$, we can immediately obtain the inequality:

$$P(A|B = true) \ \geq P(A)$$

This inequality offers an intuitive result – a positive specific observation in support of a general hypothesis will tend to increase the probability we attach to the generalisation being true. We might envisage that, with increasing observations of such positive instances, the conditional probability will tend to one. But this depends on valid assessments of P(A) and P(B). P(A), the prior probability, can be updated iteratively by Bayesian conditioning – if we see a new swan in the garden every day, then today's posterior is tomorrow's prior. Providing our 'initial' prior for A is non-zero, P(A) will tend to one if P(B) is less than one. But what is P(B)?

We may obtain some further insight into P(B) by writing it as:

$$P(B) = P(B|A = true)P(A) + P(B|A = false)(1 - P(A))$$

---

[56] Nassim Nicholas Taleb made swan colour highly fashionable in popular inductive discourse in the early twenty-first century. But the colour of swans has a significantly longer history in philosophical literature. The black swan was used by Bertrand Russell in his classic essay on induction in his book *Problems of Philosophy* published a century earlier. It also makes an appearance in R.B. Braithwaite's *Scientific Explanation* of 1953 and T.W. Hutchison's *The Significance and Basic Postulates of Economic Theory*, first published in 1938. Its original usage apparently lies with John Stuart Mill, the mid-nineteenth century English philosopher, economist and general polymath (Blaug (1980), p. 12). British actuaries can note with pride that our own Professor David Wilkie made his contribution to this esteemed tradition when he noted that Australian black swans were, in fact, partly white (Pemberton (1999), p. 192). Wilkie's observation is indicative of his general aversion to excessive philosophical navel-gazing.

Noting, as above, P(B|A) = 1 when A is true, we then obtain a new expression for the conditional probability of A:

$$P(A|B = true) = \frac{P(A)}{P(A) + P(B|A = false)(1 - P(A))}$$

This equation shows that the updating of the conditional probability relies crucially on the probability of observing a white swan when the hypothesis that all swans are white is *false*. A moment's reflection tells us why: if, for example, we knew there was no possibility of observing a white swan when the hypothesis is false (because, say, there was only one alternative hypothesis, and it was that, as of today, all swans are black), then we could deductively conclude that *the observation of a white swan shows that the white swan hypothesis is certainly true* (as, in this case, P(B|A=false) = 0 and P(A|B) = 1). Consider the other pathological case, where we are sure we will observe a white swan in the garden, even if the hypothesis is false. This could arise, say, because we *already know* all swans in Edinburgh are white, but we do not know what colour the swans are in other less explored parts of the world, such as Glasgow. In this case, P(B|A = false) = 1, and P(A|B) = P(A). This is again intuitive – in this case, observations of the colour of swans in Edinburgh provides us with no new information about the truth of the universal hypothesis, and *the conditional probability is therefore unchanged by the observation*.

In our general example, we do not have sufficient information to make the assertion that P(B|A= false) is zero or one. We cannot really say anything about P(B|A = false) from the information we have. *In the absence of a rationale for ascribing a particular value to P(B|A = false), all we can logically conclude is that observing a white swan does not reduce the probability we attach to the probability that all swans are white*. The vacuity of this conclusion is striking. It is a demonstration of Hume's problem of induction in action. It shows that the problem of induction is not just a philosopher's idle indulgence.

This example demonstrates, with the use of Bayes Theorem, that inference by mere positive enumeration, without any extra-evidential assumptions or knowledge, is indeed epistemically worthless, just as Hume argued.

Developing a Logical Definition of Probability

The philosopher's problem of induction, as well as the sheer ambition of the logical probability concept, suggests that finding a general way of quantitatively defining the rational measure of the probability of truth of a statement from a given set of (quantitative and qualitative) evidence is likely to be highly challenging. This scepticism has long been present amongst philosophers who advocate an objective approach to probability. For example, Venn's *Logic of Chance* includes the much-quoted passage:

"In every case in which we extend our inferences by induction or analogy…we have a result of which we can scarcely feel as certain as of the premises from which it was obtained. In all these cases then we are conscious of varying quantities of belief, but are the laws according to which the belief is produced and varied the same? If they cannot be reduced to one harmonious scheme, if in fact they can at best be brought about to nothing but a number of different schemes, each with its own body of laws and rules, then it is vain to endeavour to force them into one science."[57]

Nonetheless, some illustrious philosophers of the first half of the 20th century such as John Maynard Keynes, Harold Jeffreys and some of the logical positivists of the Vienna Circle such as Rudolf Carnap

[57] Venn (1888), p. 124.

pursued the idea that a given body of evidence could be used to generate a rational quantitative degree of belief in a given hypothesis. Below we consider the work of Keynes and Jeffreys, in particular. Keynes and Jeffreys were both Cambridge scholars with philosophical interests but vocations outside of academic philosophy (and indeed probability). Keynes wrote on the theory of probability in the earlier part of his career, before going on to become one of the most influential economists of the twentieth century. Jeffreys was a leading professor of geophysics and astronomy.

Keynes' 1921 *Treatise on Probability* set out his theory of probability as a system of inductive logic that could be considered as a generalisation of deductive knowledge. To Keynes, the theory of probability was the theory of rational inductive inference from limited empirical observation to partial or probable knowledge:

"That part of our knowledge which we obtain directly, supplies the premises of that part which we obtain by argument. From these premises we seek to justify some degree of rational belief about all sorts of conclusions. We do this by perceiving certain logical relations between the premises and the conclusions. The kind of rational belief which we infer in this manner is termed probable, and the logical relations, by the perception of which it is obtained, we term relations of probability."[58]

Keynes' definition of epistemic probability was based explicitly on an objective rationality:

"A proposition is not probable because we think it so. 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."[59]

As was noted above, this epistemic perspective on probability as a degree of belief was not new when Keynes was writing in the 1920s. It had formed a fundamental part of thinking on probability since the Enlightenment era and the ideas on probability that emerged then from Laplace and others. Indeed, some such as Carnap have argued that the competing frequentist definition of probability did not emerge until the 1840s, and thus probability was meant in an entirely epistemic sense prior to then[60].

Philosophically-minded mathematicians of the Victorian era also attempted to specifically develop the epistemic approach to probability, with their own nuance. For example, like Keynes, Augustus De Morgan also considered how probability could be considered as an inductive extension to deductive logic. But his departure from deduction arose not through consideration of an argument from known premises that did not logically entail a conclusion (as Keynes describes in the above quotation), but instead by assuming the premises of the argument were not known with certainty. Thus in 1847 De Morgan defined epistemic probability as "the study of the effect which partial belief of the premises produces with respect to the conclusion" [61]. This definition of epistemic probability was not significantly developed further, but it is interesting to note that it seems a representative logical depiction of scientific knowledge: that is, scientific theories are usually deductive systems developed from premises whose depiction of empirical reality is not certainly true. This topic will be discussed further in Chapter 2.

---

[58] Keynes (1921), p. 111.
[59] Keynes (1921), p.4
[60] Carnap (1950), Section 42.A.
[61] De Morgan (1847)

Sir Harold Jeffreys' *Theory of Probability* was first published in 1939 and in it he, like Keynes, set out an approach that embraced probability as the epistemic currency of inductive logic (Jeffreys also wrote another book, *Scientific Inference*, first published in 1931, which was also significantly concerned with probabilistic thinking). To Jeffreys, a probability was the 'reasonable degree of confidence that is attached to a given proposition in the presence of a specified set of data'[62]. Jeffreys, again like Keynes, argued that this probability would not be a subjective quantity that was at the mercy of the vagaries of an individual's opinion, but would be rationally and uniquely determined: "On a given set of data *p* we say that a proposition q has in relation to these data one and only one probability. If any person assigns a different probability, he is simply wrong."[63]

Whilst Keynes and Jeffreys shared a common vision of probability as the natural calculus of inductive logic, and also shared a scepticism of the frequency theory that provided the standard definition of probability in their time, their approaches differed in some important respects. Most crucially, Jeffreys took the comparability and orderability of the probabilities of different propositions as the first axiom of his probability theory[64]. In essence, *he assumed that a unique numerical value could be logically identified for the probability of anything, conditional on anything*. The theorems of the late 20th century American physicist and logical probabilist Richard Cox[65] showed that, if probabilities of propositions can be represented by real numbers (and hence are comparable and orderable as assumed by Jeffreys), then there must exist a unique set of inductive rules. Thus, if we can accept this axiom, an exciting world of inductive logic appears before us.

The property, however, that any probability can always be uniquely represented by a real number can only be an axiom or postulate of a system rather than a property that is logically derivable from more fundamental and self-evident properties. It is the crucial and essential premise necessary to enable a numerical system of inductive logic. And it is a premise that many philosophers have not been willing to accept. Keynes refused to make this assumption axiomatic of his system of probability. Instead he argued not all probabilities would be quantifiable. There would exist a subset of probabilities for which comparable statements may be possible, without quantifying either probability statement. For example, he argued it may be possible to say that a scientific hypothesis conditional on evidence *x+y* is more probable than the scientific hypothesis when only given evidence *y*. But it may not possible to say what either of these probabilities are. And then, he further argued, there would be some sets of propositions for which their probabilities were not only unquantifiable, but also incapable of comparison or ordering:

"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."[66]

The Principal of Indifference and Uninformed Priors

Keynes' and Jeffrey's inductive analyses each resolve into a form that would be familiar to Bayes and Laplace and modern subjectivists: a prior probability is combined with the results of a set of observations to obtain a posterior probability. As noted above, the special feature of the logical approach is that the prior distribution is not subjective. Rather, it is derived rationally. Today's prior is yesterday's posterior. So the ultimate logical starting point for the initial specification of the prior

[62] Jeffreys (1961), p. 20.
[63] Jeffreys (1937), p. 10.
[64] Jeffreys (1961), p. 16.
[65] Cox (1946) and Cox (1961)
[66] Keynes (1921), p. 27-8.

is its value in the case of total ignorance. A fundamental requirement of the logical approach is therefore to show *how to rationally determine the prior distribution in the case of the total absence of any information* about the proposition in question. This might at first glance sound like the simplest of all possible cases, with a self-evident answer, but it has been a crucial source of controversy in the development of the logical approach to probability.

The historical answer to the question of how to rationally specify the prior distribution in the absence of any information whatsoever was developed and applied independently by Bayes and Laplace during the final decades of the 18th century. They both asserted[67] that, in this case, the prior distribution should be the uniform distribution – in the absence of any information, all possible values are equally probable. Laplace called this the Principal of Insufficient Reason. Keynes re-named it the Principle of Indifference (a terminology which has generally stuck since) and defined it thus:

"The Principal 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 these alternatives have an equal probability."[68]

In the logical approach of Keynes or Jeffreys, the Principle of Indifference is taken to be axiomatic rather than logically-derived from more fundamental assumptions. The Principle was heavily employed by Laplace in his work on probability during the early nineteenth century (and was the basis for his (now somewhat notorious) *Law of Succession*). The Principle stood unchallenged for several decades before starting to attract some criticism in the mid-nineteenth century. The earliest critique is thought to have been produced by Leslie Ellis in 1842. The critique published by George Boole, the English mathematician and philosopher, was more expansive and influential. In 1854, Boole wrote of the Principle:

"It has been said that the principle involved in…is that of the equal distribution of our knowledge, or rather of our ignorance – the assigning to different states of things which we know nothing, and upon the very ground that we know nothing, equal degrees of probability. I apprehend, however, that this is an arbitrary method of procedure."[69]

Boole's view of the Principle of Indifference has been shared by many other leading philosophically-minded statisticians and economists of the twentieth century. To give a couple of diverse examples, Sir John Hicks, the Nobel prize-winning British economist and the frequency theorist Ludwig Von Mises have argued (separately), like Boole, that complete ignorance does not imply equal probabilities should be attached to the possible outcomes; rather, it simply means that their probabilities are not measurable[70].

As well as having a disputed logical basis, the Principle of Indifference can also be shown to produce paradoxes that arguably highlight the arbitrary nature of the attempt at equating ignorance with equiprobability. Keynes' *Treatise* provides one of the most extensive discussions of these paradoxes and attempts at their resolution.

The general form of the classic paradox of the Principle of Indifference exploits the arbitrary nature of the unit of measurement that the Principle is applied to. To take Keynes' example, suppose we

[67] Bayes may actually have been somewhat reticent in making this assertion. There is some historical speculation that it was this reticence that delayed the publication of Bayes' paper until after his death, when his friend Richard Price received Bayes' draft and immediately saw its potential value in inductive inference.
[68] Keynes (1921), p. 42.
[69] Boole (1854), p. 287.
[70] Hicks (1979), Chapter VIII.

have a quantity of substance of a known mass (say, of value 1). All we know about its volume is that it must be somewhere between 1 and 3. The Principle of Indifference would tell us that we should therefore attach equal probability to the volume of the substance being between 1 and 2 and it being between 2 and 3. Now, if we know its volume is between 1 and 3, we also know its density (mass / volume) is between 1 and 1/3. We could choose to directly apply the Principle to the density instead of the volume (why not?). The Principle tells us that it is as likely that the density is between 1/3 and 2/3 as it is between 2/3 and 1. But the median volume (2) and the median density (2/3) are not mutually consistent – the density is ½ when the volume is at its median size, not 2/3; and the volume is 3/2 when the density is at its median size, not 2. What size of volume is consistent with our knowledge that the mass is 1 and the volume is somewhere, anywhere, between 1 and 3? One approach (prior distribution for volume is uniformly distributed) says 2 and the other approach (prior distribution for density is uniformly distributed) says 3/2. And the choice of which assumption to make appears to be completely arbitrary. Moreover, this form of argument can be applied to the use of the Principle of Indifference in the case of any continuously-distributed variables (for example, instead of mass, volume and density, we could equally work with distance, speed and time).

Keynes made attempts to resolve the paradox of the Principle of Indifference, chiefly by focusing on the concept of divisibility, but his efforts were never widely accepted. A few years after Keynes' *Treatise*, Von Mises was somewhat dismissive of Keynes' efforts to solve these paradoxes, and expressed the same view as Boole 75 years earlier and Hicks 50 years later when he wrote: 'It does not occur to him [Keynes] to draw the simple conclusion that if we nothing about a thing, we cannot say anything about its probability'.[71]

Given the philosophical angst generated by the Principle of Indifference, Jeffreys was notably sanguine about its use. He believed it was a fundamental and natural outcome of his theory of probability: "When reasonable degree of belief is taken as the fundamental notion, the rule [the Principle of Indifference] is immediate….to say that the probabilities are equal is a precise way of saying we have no ground for choosing between the alternatives…it is merely the formal way of expressing ignorance."[72]

Jeffreys argued that the above paradox of the Principle of Indifference that arises with non-linear transformations of continuous variables can be addressed by subtly but fundamentally altering the interpretation of the Principle. He argued that, rather than assuming the Principle implies the 'uninformed' prior distribution must be a *uniform* distribution, instead *it implies a functional form for the uninformed prior that leaves the probabilities unchanged by these changes of variable*.

So, for continuous variables whose value must be positive, if we assume the uninformed prior distribution is a uniform distribution for the variable's *logarithm*, then the prior is invariant to the reciprocal transformation that occurs between volume and density (or, say, distance and speed), and indeed any power transformation. In the above example, this makes the median of the prior distribution for the volume the square root of 3 (approximately 1.73) instead of 2 or 3/2 as per above; and the prior median under this form of uninformed prior for the density is the reciprocal of the square root of 3 (approximately 0.577) instead of ½ or 2/3 – consistency has, by design, been achieved. Jaynes argued that this result is an example of a more general approach to specifying well-behaved uninformative priors:

[71] Von Mises (1928), p. 75.
[72] Jeffreys (1961), p. 33, 34.

"If we merely specify 'complete initial ignorance', we cannot hope to obtain any definite prior distribution, because such a statement is too vague to define any mathematically well-posed problem. We are defining this state of knowledge far more precisely if we can specify a set of operations which we recognise as transforming the problem into an equivalent one. Having found such a set of operations, the basic desideratum of consistency then places nontrivial restrictions on the form of the prior."[73]

This approach seems capable of delivering a logical solution to the paradox, albeit only by requiring careful identification of the transformations in which the prior should be invariant in order to properly reflect the circumstances left unspecified in the particular problem (and under the assumption that all probabilities are orderable and comparable). The invariance principle seems a compelling strategy to resolving the paradoxes of the Principle of Indifference. But its implications are not entirely intuitive – if we know nothing about the volume of an object other than that it is somewhere between 1 and 3, is the square root of 3 really the natural starting estimate for it? In this example, the introduction of the density variable together with the invariance requirement implies a unique solution for the uninformed prior. But why has this happened? After all, in the example we do not know anything whatsoever about the density that is not already fully expressed by what is known about the volume. Why is density more important than some other arbitrary non-linear function of volume that we could instead require to meet the invariance property? Is it because we regard density as a 'real thing', of equivalent status to volume in some ontological sense? The invariant strategy of assuming the logarithm of the variable is uniformly distributed makes the distribution invariant to any power transformation of the variable. But it is mathematically impossible to specify a form of prior that is invariant to every possible choice of transformation.

These types of question raise interesting mathematical problems for which better solutions may emerge. But, as ever with inductive inference, the fundamental contention is a logical one, and these questions have perhaps not yet been sufficiently explored by philosophers. Today, despite the support of Jeffrey's invariance principle from some notable modern quarters such as Jaynes, many philosophers still regard the Principle as, at best, a heuristic, and not as a robust starting point for a logical theory of probable knowledge.

More generally, the overall limited success in developing clear inductive rules (beyond the product and sum rules of basic probability theory) for the rational quantification of partial belief across a range of forms of evidence means that the logical theory of probability is not universally regarded as a viable philosophical or practical basis for a theory of probability. Work on the logical theory of probability has arguably failed to generate a set of practically useful general rules of rational inference – Keynes' "certain logical relations" - that can produce numerical quantities or degrees of rational belief for a broad class of inductive problems. The logical framework provides conceptual directional relationships and limits, but these generally remain far removed from the objective determination of numerical probabilities in the 'real-world'. It was this perceived failure in the work on logical probability of Keynes and the logical positivist school that provided the impetus for the development of subjective probability in the 1920s and 1930s.

The uniform prior distribution, however, is still widely used in subjectivist probability methods and arguably with some success. The logical theory needs some form of the Principle of Indifference as a logically robust postulate in order to lay claim to a measure of rational belief. Subjectivists need make no such claim of rationality for their probabilities or their more implicit use of the Principle – if

[73] Jaynes (2004), p. 381-2.

the observation of an individual's betting requirements implies a uniform distribution, so be it. To De Finetti, arguments for and against the Principle were 'meaningless and metaphysical in character'[74].

Nonetheless, in the 21st century, Bayesian methods have experienced a particular form of resurgence in scientific data analysis, and this resurgence is notable for abandoning the use of subjective priors and replacing it with the use of some form of uninformed or objective (as opposed to subjective) prior. The invariant approach pioneered by Jeffreys discussed above is one example of this approach. But a more popular contemporary approach is the use of *conjugate priors*. The basic idea of a conjugate prior distribution is that is made uninformed by making it consistent with the observational data. For example, the prior distribution may be chosen to reflect the information contained in some pre-existing hypothetical data sample, and it might be assumed that this hypothetical prior data sample has the same mean as the mean of the sample data set in consideration. This seems very far away from the original logic of subjective probability – after all, here, the prior distribution is being set with reference to the data that it is supposed to be prior to! But the uninformed prior, and the conjugate prior approach to setting it, can provide a powerful machinery of logical Bayesian inference that fits the 21st century environment of computation-heavy, assumption-light approaches to inference and prediction[75].

1.4 Some final thoughts

This brief survey of philosophical thought on probability highlights that 'probability' can be intended to mean profoundly different things, both to probability theorists and those applying statistical methods in their own fields.

To the objectivist, probability has a narrow and mathematically precise definition that relates to mass phenomena that can be repeatedly observed. The objectivist considers probability as a natural part of the domain of the physical sciences. Objectivists such as Venn recognised that this strict form of definition means that there may be a very limited field of empirical phenomena for which objective probabilities exist. But an objectivist would tend to argue that there is no point in attempting to apply a mathematical probabilistic framework to phenomena where the objective requirements (particularly of repeatability and, hence, in principle, testability) cannot be met.

The objective approach is based on the idea of probability as the *frequency* at which mass repeatable physical phenomena occur. The epistemological idea of probability as a degree of belief or measure of partial knowledge has a distinctly different starting point. Whether taking a subjective or logical approach to epistemic probability, perhaps the most fundamental question that arises therein is when and why it is reasonable to define a specific numerical probability for the degree of belief, and in what sense such a number can meaningfully exist. The question of how it is to be calculated then follows.

Historically, the logical approach arguably failed to provide sufficiently convincing answers to these questions, and the subjectivist solution that chronologically followed in the 1930s was to *observe* the degrees of belief implied by people's actions rather than attempt to *logically determine* what their rational belief ought to be. The subjective approach was given significant mathematical rigour by the Ramsay-De Finetti theorem, which showed that a very basic rationality constraint (a no-arbitrage condition for the bets that a given individual is willing to make and offer, which is called statistical coherence in probability theory) would result in subjective probabilities meeting the standard mathematical axioms of probability theory. But even the most ardent subjectivists have recognised

[74] De Finetti (1975), p. 441.
[75] Machine learning and predictive analytics is discussed further in Chapter 6.3.

that 'real-life' probabilities of uncertain events may be beyond the abstract machinery of coherence and exchangeability. To quote De Finetti,'[worked examples of quantitative updating of subjective probabilities] are useful in practice, but only as simplified representations of more complicated situations which themselves cannot be represented straghtforwardly'[76].

Whilst there is some debate about when the historical distinction between objective and epistemic probabilities fully emerged, there has been at least some epistemological aspect to probabilities since Bayes' paper of 1763 (and the approach was developed and more widely disseminated by the probability work of Laplace at the start of the 19th century). This established the prior / evidence / posterior structure of inductive inference. There has been an explicitly objectivist train of thought in probability theory since at least the time of George Boole's 1854 book *An Investigation into the Laws of Thought.* This school of thought has argued that there are many situations where there is no rigorous logical basis for assuming a prior probability of a given specific numerical quantity. In such cases, the updated posterior probability must always contain an arbitrary constant (the prior probability). Boole argued that a lack of intellectual clarity or honesty about this could over-state the degree of precision and the quality of knowledge that we can possess about an uncertain world. He protested against "[numerical probabilities'] intrusion into realms in which conjecture is the only basis of inference."[77]

Boole's argument, carried forward over the years by Venn, Fisher and many others, perhaps remains the fundamental protest of the objectivist against the subjectivist's approach to probability: that it is a futile attempt to measure the immeasurable, and that conducting our affairs under the illusion that we have obtained a form of knowledge that is stronger than it really is can be a dangerous error.

Whilst objectivist and subjectivist perspectives on the meaning of probability may appear irreconcilably at odds, many leading philosophers and scientists over the last one hundred years or so have advocated a pluralist view which asserts that these different definitions of probability may apply in distinct domains. Advocates of this approach include a diverse set of intellectual luminaries such as Henri Poincare[78], Frank Ramsay, Sir John Hicks[79], Rudolf Carnap[80], Ian Hacking[81] and Donald Gillies[82]. This approach recognises that the word 'probability' can simply be used to mean different things in different places. It is also notable that the surge in popularity of Bayesian methods of recent decades has been mainly characterised by the use of methods that use 'uninformative' priors. The idea of using priors to reflect qualitative subjective information has been largely abandoned in today's data science algorithms. Instead, approaches such as empirical and objective Bayesian methods combine both frequentist and logical elements.

So perhaps we can opt to recognise that there will be some situations where objective probabilities are natural (for example, in describing the behaviour of seemingly irreducibly statistical physical phenomena such as the radioactive decay of uranium isotopes) whilst simultaneously recognising that there other situations where (inter)subjective or logical probabilities are applicable (for

---

[76] De Finetti (1975), p. 432.
[77] Boole (1854), p. 286.
[78] Poincare (1902), Chapter XI.
[79] Hicks (1979)
[80] Carnap (1950)
[81] Hacking (1965), p. 226-7.
[82] Gillies (2000), Chapter 9.

example, in the probabilistic evaluation of unique, singular phenomena, such as the behaviour of next year's inflation rate).

Amongst all this philosophical debate, one area of consensus appears to be identifiable amongst objectivists from Venn to Popper and subjectivists such as De Finetti: the strict demands of the objective definitions of probability mean that objective probability should have little applicability in the social sciences (objectivists such as Venn and Popper argued that social sciences are inherently non-stationary, rendering objective probability inapplicable; whereas subjectivists such as De Finetti argued objective probability is inapplicable everywhere). Yet objective probability theory (and its statistical inferential techniques) are applied all the time across a range of social sciences, including a range of modelling fields (financial and mortality) in actuarial science (though Bayesian methods have emerged as significant alternative statistical approaches in social science fields such as econometrics since the early 1970s[83]). Moreover, if the reason for rejecting the use of objective probability in the social sciences is because empirical data in the social sciences tends to be non-stationary and not necessarily independent, then this may also preclude the application of many standard subjective probability modelling techniques – recall that Bayesian conditioning depends on the sample data having the exchangeability property, and that this in turn is equivalent to objective independence. No matter how we define our concept of probability, statistical forms of inference will depend on observations that are, in some sense, well-behaved and representative of the unobserved. The problem of induction otherwise has no solution. We cannot learn from past experience unless there is reason to believe the experience will have some well-understood relation to the future. We will return to this topic in the later chapter dedicated to probability and methodology in the social sciences.

Before we tackle the special ambiguities of the social sciences, first we will consider what the above discussion of probability and inference means for methodology in the natural sciences. In particular, can probability theory be usefully employed to measure the degree of belief in a scientific hypothesis? This is a question that will be considered as part of a broader survey of the philosophy of the scientific method in Chapter 2.

[83] See, for example, Zellner (1980).

2. Probability and methodology in the natural sciences

"Probability is not the subject-matter of a branch of science; science is a branch of the subject-matter of probability. To ignore probability is to ignore the problem of scientific inference and to deprive science of its chief reason for existence." *Sir Harold Jeffreys, Scientific Inference (3rd edition), p.219, 1937.*

In antiquity, science was mainly concerned with the development of certain knowledge by the logical deduction of true statements from premises that were taken to be self-evidently true. Euclidean geometry, which derived an impressive array of mathematical results from only a handful of axioms, is the canonical example. In this setting, empirical observation merely plays the role of confirming knowledge that has already been deductively proven.

This view of science was radically altered during the 200 or so-year period from the mid-16th century to late-18th century that spanned the scientific revolution and the philosophical developments of the Enlightenment. Francis Bacon's writings on an empirically-based scientific method and Hume's work on causation and the problem of induction were major contributions, amongst many others, that helped to provide the basis for the development of the modern scientific method and its philosophical study. The scientific method that has been developed and implemented over the past 300 years has delivered staggering progress and success in the fields of the natural sciences.

Today, science is concerned with describing empirical observations in a way that permits wider and simpler generalisations about the behaviour of nature. Such generalisations can deliver accurate predictions about the behaviour of future (or otherwise unobserved) instances of phenomena. This basic statement suggests that the challenge of building scientific knowledge is fundamentally one of inductive inference – that is, of using knowledge of the observed to infer properties or behaviours of the unobserved. The problem of induction tells us that inductive inference requires principles, assumptions and techniques beyond those required in deductive logic. Scientific knowledge, in contradistinction to the certain knowledge of deductive logic, must always be partial, uncertain, fallible and, at best, *probable*. It is in this sense that the Jeffreys quote above refers to science as a branch of the subject matter of probability.

In the context of the discussions of Chapter 1, it is therefore natural to ask: does probability have a role to play in measuring the strength of scientific knowledge? Does this perspective provide insights into what features are desirable for the scientific method? Do objective and epistemological philosophies of probability provide different answers to these questions? This chapter provides a brief overview of philosophical thought on the methodology of science, with a particular focus on the role of probability and the various answers to these questions that they imply.

Science is a term that, today more than ever, can cover a very broad church of subjects. This chapter will focus on methodology as it generally pertains to the natural sciences (physics, chemistry, biology, etc.). Chapter 3 will then extend this discussion to consider the special methodological difficulties that arise in the social sciences (such as economics, sociology and psychology).

In the discussion of probability and methodology in the natural sciences that follows in the first two parts of this chapter, we will consider two distinct schools of thought in the philosophy of scientific method (falsificationism and Bayesianism). These two perspectives can be closely related to the two distinct philosophical views of probability that were discussed in Chapter 1 (objective probability and epistemic probability). Perhaps unsurprisingly given the potential link between philosophy of probability and the philosophy of scientific knowledge, we will find that two of the leading proponents of the competing probability theories (who are already known to us from Chapter 1) also contributed very significantly to thinking on scientific method – in particular we will consider the work of Sir Karl Popper in developing falsificationism as a theory of scientific method that is consistent with objective probability; and the work of Sir Harold Jeffreys in applying an epistemological probability theory to develop his contrasting view of how scientific knowledge is discovered and tested.

Following the discussions of falsificationsim and Bayesianism of Chapters 2.1 and 2.2, 2.3-2.5 will then move beyond the topics that are most closely related to probability theory, and briefly consider other key topics of philosophy of science that may still have implications for the methodology of actuarial science.

But first it may be useful to note a couple of basic properties that tend to be found in any (orthodox) perspective on scientific method. The first and most fundamental of these is that science is inherently *empirical*. Scientific theories stand or fall on the extent to which they explain and / or predict behaviour that is seen in the 'real' natural world. This, in turn, demands a need for some sort of methodological postulate about the *uniformity of nature* – if there are not natural regularities that exist across space and time, scientific generalisation from empirical observation cannot hope to succeed. Even Popper, who tried to reject inductive inference and its associations wherever possible, conceded this point: "Scientific method presupposes the immutability of natural processes, or the 'principle of the uniformity of nature'."[84]

Second, since the medieval writings of William of Ockham and his famous razor, if not before, the concept of *simplicity* has been viewed as an intrinsically desirable property of scientific theories. Simplicity means explaining more from less. Karl Pearson provided a concise philosophical expression of Ockham's razor:

"The wider the range of phenomena embraced, and the simpler the statement of law, the more nearly we consider that he has reached a fundamental law of nature." [85]

Putting differences aside on what philosophers mean by a fundamental law of nature, philosophers of science of most hues would agree with the proposition that simplicity is a basic desideratum of science. Poincare[86], Popper[87], Pearson and Jeffreys[88] and many other leading figures in the philosophy of science have emphasised that it is the essence of science to describe as much natural phenomena as possible in the simplest of terms.

We now turn our attention to the specifics of the two philosophical schools of falsificationism and Bayesianism and how they may be applied to the logic of the scientific method.

[84] Popper (1959), p. 250.
[85] Pearson (1911), p. 96.
[86] Poincare (1902), Chapter IX.
[87] Popper (1959a), Chapter 7.
[88] Jeffreys (1937), Chapter IV.

2.1 Falsificationism – an objectivist's approach to scientific inference

"The strongest argument in favour of the truth of a statement is the absence or impossibility of a demonstration of its falsehood."[89] *Karl Pearson, The Grammar of Science, 1911*.

Falsificationism was first introduced as a well-developed philosophy of the scientific method by Karl Popper some eighty-five years ago in his *Logic of Scientific Discovery*[90]. The book must stand as one of the twentieth century's most influential books on the philosophy of science. Popper's *Logic* was first published when he was only 32, and he published extensively for many of the following decades, but it can be regarded as his *locus classicus*. There must be very few fields of methodological study in the second half of the twentieth century, across both the natural and social sciences, which do not explicitly consider how the methodology of their discipline stacks up against the tenets of Popper's falsificationism. Today, perhaps with some extensions and / or caveats that will be discussed later in this chapter, Popper's falsificationism is still regarded by many as a model or benchmark for an idealised logic of the scientific method. Like most good ideas, some historical anticipations can be identified in the works of earlier important thinkers, as illustrated by Karl Pearson' quote above. Popper did not invent falsificationism, but he advocated and developed it as a form of inductive logic that he believed was at the core of a scientific method that was up to the task of forming scientific knowledge with meaningful empirical content.

Popper, the Vienna Circle, Probability, Induction and Scientific Knowledge

The decade between the mid-nineteen twenties and mid-thirties was a very fertile period for the philosophy of science. It was during this period that the Vienna Circle of logical positivists emerged as a group of thinkers that would have major influence on the path philosophy of science would take over much of the rest of the twentieth century. The logical positivists' doctrine was notable for its empiricism and its view of metaphysics as being devoid of scientific meaning. Popper shared some (but certainly not all) of these philosophical values, was born in Vienna and attended the University of Vienna. But he was not a member of the Vienna Circle and would have rejected having the label of logical positivist applied to him.

The Vienna Circle, at least in its early period, argued that a synthetic[91] statement only had meaning if it was empirically *verifiable* – that is, if empirical evidence could conclusively confirm the statement to be true. This is a very high epistemic bar, as it demands certainty. In particular, it would seem impossible to empirically verify the truth of any universal generalisation (as such a generalisation applies to an infinite number of instances, making it impossible to empirically verify all applicable cases have the asserted property). Most scientific hypotheses imply a form of universal generalisation (i.e. a hypothesis states that, in given specified conditions, specified forms of behaviour between some phenomena will always occur). Over time, the logical positivists recognised that their requirement that meaningfulness required empirical verification was unrealistic: the (hugely successful) scientific knowledge that had been developed over the previous two hundred years could not meet this benchmark of empirical verification. Such knowledge was fallible, and not verifiable. But it surely wasn't devoid of meaning.

The logical positivists' solution to this quandary was to attempt to develop a logic of inductive probability (Rudolf Carnap was especially notable in this field): if absolute empirical verification was

[89] Pearson (1911), p. 56.

[90] Popper (1959a). Note the book was first published in German in 1934, and the English edition did to appear until 1959.

[91] 'Synthetic' is used here in the Kantian sense of referring to a proposition that relates meaning to some sort of aspect of the world that is not wholly defined by the premises of the proposition.

not attainable for scientific knowledge, then the next best thing would be a logical enumeration of the degree to which the empirical evidence 'partially verified' it. We saw in Chapter 1 that Popper took a different view of the meaning of probability. To Popper, probability was a property that was strictly related to mass phenomena generated under repeatable conditions. He did not believe there was a role for probability as a measure of the degree of belief in the truth of a given synthetic statement, including a scientific hypothesis.

Popper was profoundly influenced by Hume's problem of induction. A deep scepticism of the value of 'naïve' inductive inference by mere enumeration of positive instances is present throughout much of his work. Popper's views on probability and inductive inference might appear to place him in a philosophical cul-de-sac – inductive inference from empirical evidence could neither verify the truth of a scientific hypothesis nor quantify the degree of belief which should be attached to the hypothesis. Did this imply there was no way for science to obtain knowledge? If the testing of scientific theories against empirical observation was a futile inductive process empty of epistemic content, how could science exist and successfully develop in the way that it evidently historically had done?

The Falsificationist Solution

Popper's solution was to argue that empirical testing of scientific theories need not mean only using empirical evidence to build a *positive inductive* case for the *truth* of the theory. Empirical evidence could also be used to *deductively* demonstrate that the theory had *not been proven to be false*. Recall the colourful swan example of Chapter 1.3 where we found that any number of positive observations need not necessarily result in a probability of 1 being attached to the hypothesis that all swans are white. Yet one single observation of a non-white swan reduces the probability to 0 – and this arises as a piece of deductive logic, not inductive inference. This type of deductive argument, where a negative instance shows a universal law must be untrue, is an ancient part of logic and even has a formal Latin name – *modus tollens*.

Popper's philosophy of science used this logical asymmetry to argue that epistemically valuable scientific knowledge could be derived from empirical observation despite his belief that the simple enumeration of successes by induction was of no epistemic value. This implied empirical observation should be used to empirically test a scientific hypothesis in ways that had the potential to *falsify* it. And the more tests the hypothesis survived, the greater the strength of belief in the theory (though this strength could not be quantified as an epistemic probability and could never reach certainty).

Note that the logic of falsificationism does not render scientific knowledge into deductive knowledge: in falsificationism, the theory is not deduced to be true or false. Rather, falsificationism is a form of inductive logic which Popper argued had (much) more epistemic value than the mere enumeration of positive instances. And there is nothing about the logic of falsificationism that would necessarily be disagreeable to a logical or subjectivist perspective of probability. The key differentiator between Popper and the logical or subjectivist perspective in this respect is that Popper rejected the use of quantitative epistemic probability as a measure of degree of belief in the scientific hypothesis (we will discuss further below what he proposed using instead).

To Popper, falsificationism was so essential to the scientific method that it represented the demarcation between science and metaphysics (or anything non-scientific) - if a theory could not be subjected to tests that could produce a falsifying result when the theory was not true, then the theory could not be considered as belonging to science. This can be seen as Popper's response to the logical positivists' attempt to demarcate scientific and non-scientific knowledge by empirical verification.

The Falsificationist Scientific Method

The falsificationist philosophy implies a continual cycle of scientific progress where a new theory supersedes the last in the generality, accuracy or reliability of its implications. Below we will find that philosophers of science have identified several logical challenges and real-life complications in relation to the basic falsificationist model. But before discussing those, we first briefly summarise the key steps of this idealised model of scientific progress:

1. Collect and classify empirical data through experiment and / or observation. Conjecture a theory that fits this data and which is consistent with background knowledge and other accepted scientific theories.

An infinite number of theories may fit a finite set of empirical statements. For reasons that will become clearer below, 'good' properties of a scientific theory according to falsificationist principles will include simplicity and empirical testability. The development of a new scientific hypothesis was viewed by Popper as a form of creative or inventive process. A theory is a human construction for description and prediction of empirical phenomena. The falsificationist does not stipulate particular logical or methodological constraints on this creative process. The theory could be inspired by metaphysical ideals or come from other non-scientific inspiration.

In this initial stage of the process of theory development, it may be difficult to say which comes first: the empirical observation or the theory. At one end of this spectrum might be the Baconian view of collecting as much empirical data as possible before formulating a theory that fits these facts (Bacon's famous "countless grapes, ripe and in season"[92]). At the other end of the spectrum, theories may be conjectured that are motivated by primarily metaphysical ideas rather than already-identified empirical patterns (Democritus' theory of the atom, for example). Popper had a somewhat romantic notion of scientific theory creation that arguably sat closer to this end of the spectrum than to Bacon's indiscriminate observing: "Bold ideas, unjustified anticipations, and speculative thought, are our only means for interpreting nature."[93]

2. Logically deduce what the conjectured hypothesis implies for the behaviour of observable phenomena (beyond the observations considered above).

This part of the process is fundamental to the falsificationist philosophy (and any empirical philosophy of science). It determines how the hypothesis can be empirically tested. For a falsificationist, the critical output from this stage of the process is an unambiguous determination of what the hypothesis implies *cannot* empirically occur (what Popper referred to as the theory's 'potential falsifiers'). The more that the theory says cannot happen, the more testable and falsifiable the theory is, and the greater is the theory's potential empirical content:

"If the class of potential falsifiers of one theory is 'larger' than that of another, there will be more opportunities for the first theory to be refuted by experience; thus compared with the second theory, the first theory may be said to be 'falsifiable in a higher degree'. This also means that the first theory says more about the world of experience than the second."[94]

Popper argued that simple theories rule out more than complex theories do. This therefore provides the falsificationist's rationale for simpler hypotheses being preferred to complex theories – they are, all other things being equal, 'falsifiable in a higher degree'.

---

[92] Bacon (1620), p. 123.
[93] Popper (1959), p. 280.
[94] Popper (1959), p. 96.

This is all consistent with the idea that science's aim is to explain as much as possible from the fewest and simplest assumptions. The most successful and important scientific theories tend to have implications for a very wide array of phenomena, and it is the success of the theory's predictions relative to the empirical behaviour of such a wide array of phenomena that provides the theories' high empirical content. Quantum theory, Newtonian mechanics and the kinetic-molecular theory are some historical examples of scientific theories which could claim to be falsifiable to a very high degree due to the testable implications that they have for a very diverse range of physical phenomena.

3. Test the accuracy of the theory's new predictions by collecting relevant data on the predicted phenomena through experiment or observation.

The theory will be said to be 'corroborated' when the empirical data of the behaviour of the new phenomena matches the predictions of the theory. The theory is never verified as true. As noted above, Popper's philosophy of objective probability rules out attaching a probability to the theory being true. He argued that a theory could, however, be *corroborated to different degrees*:

"The whole problem of the probability of a hypothesis is misconceived. Instead of discussing the 'probability' of a hypothesis we should try to assess what tests, trials, it has withstood; that is, we should try to assess how far it has been able to prove its fitness to survive by standing up to tests. In brief, we should try to assess how far it has been 'corroborated'."[95]

It is fundamental to Popper's philosophy that the empirical corroboration of a hypothesis does not have a binary state. A hypothesis that passes new potentially falsifying tests could be said to be corroborated to a greater degree than it was before. This 'degree of corroboration' sounds fundamentally similar to the Bayesian idea of determining the posterior probability of the hypothesis (discussed further below in Chapter 2.2). And it clearly behaves in some similar ways. For example, Popper even notes that "[a theory's] degree of corroboration will increase with the number of its corroborating instances."[96] He also states:

"It is not so much the number of corroborating instances which determines the degree of corroboration as the severity of the various tests to which the hypothesis in question can be, and has been, subjected."[97]

Again, none of this is inconsistent with the behaviour of a Bayesian posterior probability. The fundamental difference between the Bayesian posterior probability of the truth of a scientific hypothesis and the Popperian degree of corroboration of a scientific hypothesis is that, although it appears it will be possible to at least sometimes rank the degree of corroboration that has been achieved by two sets of tests, the degree of corroboration will generally not be capable of quantification and will not correspond to an epistemic probability of truth:

"The degree of corroboration of two statements may not be comparable in all cases…we cannot define a numerically calculable degree of corroboration, but can speak only roughly…"[98]

[95] Popper (1959), p. 248.
[96] Popper (1959), p. 268.
[97] Popper (1959), p. 266.
[98] Popper (1959), p. 266.

So, positive empirical evidence in support of the theory will increase its degree of corroboration and negative evidence will falsify the theory. Popper, however, notes that the negative evidence must be a 'reproducible effect'[99] to qualify as a falsifier of the hypothesis (we will return to this point below).

4. When a theory has been falsified, the theory may be rejected altogether, or it may be amended or extended by an auxiliary hypothesis. Such a change must not be *ad-hoc* (i.e. merely fit better to the considered empirical data in order to now pass the given test) but must provide new content (new predictions by means of logical deduction from the postulates of the new hypothesis) which can be corroborated as per above.

Again quoting from Popper's *Logic*: "As regards auxiliary hypotheses, we propose to lay down the rule that only those are acceptable whose introduction does not diminish the degree of falsifiability or testability of the system in question, but, on the contrary, increases it."[100]

This is another way in which the falsificationist method encourages simplicity: complexity can only be added if it creates new predictions which can be corroborated by new evidence. This also demonstrates how the falsificationist methodology should create a scientific process that is continually creating new forms of higher scientific understanding. Every new theory should be comparably better than the old theory:

"A theory which has been well corroborated can only be superseded by one of a higher level of universality, that is, by a theory which is better testable and which, in addition, contains the old, well corroborated theory – or at least a good approximation to it."[101]

In the Popperian scientific method, the above basic steps form a cycle of never-ending scientific progress. Whilst the method above has been cast as part of falsificationist logic, it forms the essence of the positive scientific method more generally, irrespective of whether we subscribe to the view that empirical testing can deliver strict falsification of a scientific theory or whether we take degree of corroboration or posterior probability of truth as our measure of empirical content. It is sometimes referred to as the hypothetico-deductive method. It is clearly a simplified model of how scientific knowledge develops 'in the real world'. We now turn to a couple of important philosophical and logical complications that the falsificationist model of scientific logic gives rise to (and some potential sociological complications will be briefly discussed in Chapter 2.5).

Falsification of Systems of Theories: Duhem-Quine and Braithwaites' Hierarchical Structure of Theory

The logic of falsificationism (in its basic or 'naïve' form) relies on the ability to conduct a crucial experiment or test that will, in the event that the hypothesis is actually false, produce results that single-handedly lead by logical deduction to the unequivocal conclusion that the hypothesis is indeed false. The *Duhem-Quine thesis* represents a challenge to the logically deductive nature of falsificationism. It states that it is impossible for such a crucial experiment to exist.

This may, at first glance, seem an odd proposition, given the very direct and straightforward nature of the *modus tollens* logic of falsificationism. The thesis, however, argues that a theoretically-predicted experimental result will always rely on a number of auxiliary assumptions and / or other hypotheses as well as the hypothesis that is intended to be tested. That is, the logical deductions that have been made in order to make the observational prediction will always involve a form of *deductive system*, rather than only a single hypothesis.

---

[99] Popper (1959), p. 66.
[100] Popper (1959), p. 62.
[101] Popper (1959), p. 276.

This implies an experiment can therefore only test the deductive system as a whole – that is, it can only test a number of hypotheses and assumptions simultaneously. If and when an experiment produces results that are impossible according to the deductive system, we can deduce that some element of the logical system is false. But it is not possible to say *which* of the hypotheses or other auxiliary assumptions of the deductive system should be rejected.

The essence of this thesis was first developed by Pierre Duhem, the French scientist and philosopher, in 1904-5[102] (it was further refined by the American philosopher of science Willard Quine in the 1950s). Its origin therefore pre-dates Popper's work on falsificationism by decades and throws the whole logic of deductive falsification through experimental observation into question. However, in theory, the logic of falsification need not rely on a *single* crucial experiment. It may be possible for the scientist to design a system of repeatable experiments or collate a series of different observations that can allow the various hypotheses and assumptions to be analysed and isolated in turn. Nonetheless, the Duhem-Quine thesis suggests that the falsificationist method described above can only be viewed as an idealisation of a workable scientific method. The complexities of the real world will mean that the professional judgement (or hunch) of the scientist may be necessary in the appraisal of scientific knowledge and, in particular, when deciding which parts of a system of hypotheses should be rejected.

In the 1940s and '50s, a new generation of philosophers of science took over the mantle of positivism from the Vienna Circle. This generation became known as the logical empiricists, and notable members included Richard Braithwaite, Ernest Nagel and Carl Hempel, all of whom are discussed further below. Logical empiricists were interested in, amongst other things, the logical structure of scientific theories and its implications for evidence, testing and confirmation. Their work had potentially important implications for the logic of falsificationism in the presence of the Duhem-Quine thesis.

Braithwaite argued that a scientific theory could be considered as a *hierarchal* deductive system of hypotheses and postulates[103]. The system may contain higher-level hypotheses that include theoretical, non-observable terms that cannot be directly empirically tested. From these higher-level hypotheses, lower-level hypotheses may be deduced that consist entirely of observable terms that do have empirically testable implications. The structure may consist of any number of levels and hypotheses.

Even in the case of the lower-level hypotheses, it is important to note that a hypothesis can still be considered as a formal, abstract, self-contained piece of deductive logic absent of any necessary empirical content. (The philosophical literature abounds with conflicting jargon for a theory when it is considered as an abstract logical system divorced from empirical interpretation: 'pure theory', 'model', 'abstract law', 'analytical system' etc.). It will therefore be necessary to specify how the terms of the theory correspond to empirical counterparts. This gives the theory empirical content, making it a scientific hypothesis (or 'applied theory' or 'synthetic').

Furthermore, where the empirically observable objects of the theory are unambiguous, there may often still be a 'gap' between the theory's empirical implications and what can be easily observed in empirical reality. That is, the lower-level scientific hypotheses that imply specific behaviour for observable phenomena may require further specifications or assumptions in order to fully articulate the testable physical implications of the theory under a set of given conditions. In contemporary

[102] Duhem (1904-5)
[103] Braithwaite (1953)

philosophy of science, this gives rise to one of the more common usages of the term 'model'. For example, the theory of classical mechanics can give rise to a model of a pendulum which can then yield empirically testable results. The pendulum model is intended to apply the mechanical theory to some empirically specific phenomena, but gaps may remain due to real-life observational complications such as the effect of air friction. So, in this sense the model is an idealisation of reality. It may be possible to develop approximations or use other theory to fill in or correct for the gaps in the model, thus making the model an ever-improving representation of reality (assuming its underlying theory 'works' in this applied domain). The role of scientific models and their interface with theory and reality has been a major topic in the philosophy of science since the 1980s[104].

According to Braithwaite's structure, a single given hypothesis may be the product of several higher-level hypotheses, and these may lead to a large number of lower-level testable hypotheses. The same high-level hypotheses may therefore be 'embedded' in multiple theories across the given scientific field (or even across various scientific fields in the cases of the most general and important scientific theories). The empirical corroboration of a given lower-level hypothesis will provide indirect empirical support for all the higher-level hypothesis from which it is deduced. By increasing the empirical support these higher-level hypotheses, this in turn provides further empirical support for the other lower-level hypotheses that have been deduced from them.

When faced with a falsifying observational result for the lower-level hypothesis, the scientist now has a choice of which hypotheses to reject – the lower-level hypothesis with the directly testable empirical consequences has been falsified; if this hypothesis has been jointly deduced from higher-level hypotheses in a logically definitive way, then at least one of these higher-level hypotheses must also be false. If several higher-level hypotheses have been used to deduce the lower-level hypothesis, the question then arises of which higher-level hypothesis to consider falsified?

The scientist will likely be inclined to make the choice that is the least disruptive for the overall deductive system of the science. This means that, in a sense, some (highest-level) hypotheses of the science will essentially be unfalsifiable with respect to that observation at that time. In this case, the hypothesis has become what the philosopher Arthur Pap called 'functionally a priori'[105]. Braithwaite argued that this *a priori* status would only be temporary[106]. If indeed the higher-level hypotheses of the system were no longer valid premises for the phenomena under consideration, the emergence of further falsifying results for the system would encourage a wider revision of the system, and ultimately a reconsideration of some of its highest-level hypotheses.

The 'unfalsifiability' of any single hypothesis by any single experiment or observation that arises in Braithwaite's hierarchical structure is, of course, well-aligned with the Duhem-Quine thesis discussed above. Braithwaite's hierarchical structure can be thought of as a useful model for articulating the point made by the thesis. Quine's original and major contribution to the thesis, the essay *Two Dogmas of Empiricism* in his book *From a Logical Point of View*[107], was published in the same year (1953) as Braithwaite's *Scientific Explanation*. Quine argued that 'the unit of empirical significance is the whole of science'[108]. This became known as a *confirmational holism*. Where Braithwaite described a hierarchy of hypotheses, Quine wrote of a man-made fabric which impinges

---

[104] See, for example, Fraasen (1980), Cartwright (1983), Giere (1988), Morgan and Morrison (1999) and Suarez (2009).
[105] Pap (1946)
[106] Braithwaite (1953), p. 112.
[107] Quine (1953)
[108] Quine (1953), p. 42.

on experience only along the edges. The essential point is the same: the total field of scientific knowledge is underdetermined, and much of this knowledge can only be subjected to empirical testing indirectly. Quine argued that this meant that the conventional philosophical distinction between theory and empirical application – Kant's analytic and synthetic; Hume's ideas and facts; Popper's metaphysics and science; Weber's abstract and concrete – was really a difference in degree rather than a difference in kind. He even argued that the epistemic validity of pure mathematics was based on its use in empirical science (that is, in Braithwaite's terminology, as the highest of higher-level hypotheses)[109].

So far, this discussion is rather abstract. Let us now illustrate how the logic of falsificationism and the Duhem-Quine thesis can apply to a famous example from the history of science. The planet Uranus was discovered by telescope in 1781 by William Herschel. Herschel initially thought he had found a comet, but astronomers established over the following few years that the object he had identified was in fact orbiting the sun and it was subsequently established that it was actually the seventh planet of the solar system. Observations of the path of Uranus over the following few decades, however, showed that its path was not exactly as predicted by Newton's laws of gravity and mechanics. The logic of Popperian falsification would suggest that this observation could be interpreted as a falsification of Newton's laws and, hence, at least one of the higher-level hypotheses of Newtonian mechanics should be rejected and a new theory of gravitation and mechanics would be required to replace it. But this isn't what happened. Scientists' conviction in Newtonian mechanics was sufficiently strong that they preferred to interpret this evidence differently – the Newtonian prediction of the path of Uranus relied on the auxiliary assumption that there was not another significant mass nearby whose gravitational force could act upon the path of Uranus. Scientists believed that it was much more likely that this auxiliary assumption was false than that the laws of Newtonian mechanics were false. And this, in turn, was a judgement that was based on the degree to which Newtonian laws had already been corroborated in other similar settings, and the feasibility that another as-yet unidentified object was in the vicinity of Uranus' path.

Astronomers were able to infer the mass and orbit of the as-yet unidentified planet that would make the laws of Newtonian mechanics fit with the observed path of Uranus. This hypothesised planet, Neptune, was duly observed by telescope in 1846, and its observed path was indeed as predicted by the Newtonian calculations. So, what initially looked like evidence falsifying Newtonian mechanics actually provided it with an impressively powerful empirical corroboration (albeit some 65 years after the initial potentially falsifying observation).

This story has an interesting twist in the tail. Around the time of the discovery of Neptune, astronomers also identified that Mercury's behaviour was slightly anomalous relative to its Newtonian predictions (technically, its perihelion was advancing slightly faster than predicted). Using the same logic as that which led to the prediction and discovery of Neptune, astronomers hypothesised that there was another planet, yet to be observed, in an orbit closer to the sun than Mercury. Confidence in this tried and tested logic was sufficiently high for the hypothesised planet to be given a name - it was called Vulcan. But the seemingly inevitable visual observation of the planet Vulcan never came.

In 1905, Einstein published his general theory of relativity, which essentially superseded Newtonian mechanics (or, at least, showed that Newton's inverse square law was only a very good approximation to the relationship between distance and gravitational force implied by Einstein's theory). The theory of relativity predicted the behaviour of Mercury's perihelion very accurately,

[109] For an accessible discussion of Quine's philosophy of mathematics, see Linnebo (2017), Chapter 6.

without any need for the planet Vulcan. This falsified the hypothesis that Vulcan existed. It also provided a logically robust falsification of Newtonian mechanics. Scientists, however, still use Newtonian mechanics very widely, and it is still taught to all students of physics. The general theory of relativity did falsify Newtonian mechanics, but it also clarified how accurate and reliable Newtonian mechanics could be as an approximation to relativity within a particular domain of parameter values for velocity, mass and distance.

Falsification and Probabilistic Hypotheses

Popper's motivation for the development of the falsificationist methodology was fundamentally driven by his scepticism about the epistemic value of inductive inference by mere enumeration of positive instances. As noted above, falsificationism avoids a reliance on this form of induction by exploiting the deductive consequence of the negative observation: whilst an infinity of positive observations is insufficient to verify the truth of a universal generalisation, one negative observation can be sufficient to demonstrate it is false (the Duhem-Quine thesis notwithstanding).

However, there is a particular form of scientific hypothesis where a negative observation cannot logically falsify the theory, even in the absence of any Duhem-Quine considerations. Under Popper's criterion for demarcation between science and metaphysics, this may appear to be a contradiction: we noted that, according to Popper, the falsifiability of a theory was necessary for it to be considered as a scientific theory. But there was an interesting and important exception that was impossible for Popper to ignore, especially during the era in which he wrote the *Logic of Scientific Discovery*.

This exception is where the scientific theory includes a probability statement. Or, more specifically, where the deductive consequences of the theory are probabilistic. In this case, a universal generalisation is no longer of the form 'all A are B', but instead something like, 'some specified proportion of A are expected to B'. It was impossible for Popper to argue such forms of theory were metaphysics (or anything else) rather than science for one very simple reason: the most important and successful scientific theory of his time, quantum theory, used probability statements fundamentally and irreducibly. And the relevance of probability statements in science has arguably only grown since the time of publication of Popper's *Logic*, for at least two distinct reasons: it has become apparent that the irreducibly probabilistic nature of quantum theory is not a passing fad, but is likely to be a feature of our understanding of the fundamentals of natural science for a long time to come; second, the social sciences have developed significantly in scale and breadth over the last one hundred years, and, as we shall see in Chapter 3, probability statements tend to play a major role in the fields of social science.

When a scientific theory uses probability statements and produces deductive implications that are probabilistic, it may not be logically possible to falsify it. The theory may imply a non-zero probability for literally any given observation, in which case no observation can lead to the logical deduction that the theory is false (unless the observation falls outside the bounds of the specified probability distribution for the outcome, but as these bounds may be infinite, that does not always help).

Popper fully recognised this issue and argued that 'practical falsification' could be considered to have been attained when the probability of the observation occurring according to the theory was extremely low. In essence, he advocated the use of what a statistician would recognise as a hypothesis test where the null hypothesis is that the scientific theory is true. He argued that this approach fitted with his criteria for falsification, as these criteria included the requirement for the falsifying observation to be a 'reproducible effect'. Such improbable observations would not be regularly reproduced in further observations if the theory was indeed correct:

"The rule that extreme improbabilities have to be neglected agrees with the demand for scientific objectivity….I do not deny the possibility that improbable events might occur…What I do assert is that such occurrences would not be physical effects, because, on account of their immense improbability, *they are not reproducible at will*."[110]

Whilst this is doubtless the only practical solution, the presence of probabilistic statements creates particular methodological considerations for falsificationism. Falsificationism's greatest asset – the uncompromising clarity of its deductive logic – is impaired by the presence of probability statements. With a deterministic theory, a falsifying result is final, and there is no logical possibility of the rejection of the theory being wrong other than through Duhem-Quine considerations. In the case of probabilistic theories or statistical generalisations, a falsifying conclusion must always be provisional and subject to 'cancellation' – it is logically possible that further evidence will emerge that will show that the previous falsifying evidence was merely a statistical outlier. When such new evidence emerges, the only logical outcome is for the scientist to change their mind and conclude that the theory is not falsified after all.

The extent to which the above 'cancellation' of a falsification conclusion is likely to occur does of course depend on the specific circumstances of the empirical study, and the statistical strength of the initial falsification. This brings us to the next methodological complication with empirically confirming probabilistic theories: the 'extreme improbabilities' that Popper mentions must be quantified in some way. Specifically, we must choose a degree of improbability as a criterion for 'practical falsification'. As in standard hypothesis testing theory (see Chapter 1.2), this choice involves a trade-off between the probability of false negatives (i.e. concluding the theory has been falsified when it is true; a Type I error in hypothesis testing jargon) and the probability of false positives (i.e. concluding the theory has not been falsified when it is false; a Type II error). The 'optimal' choice here must have regard to the costs and consequences of these errors. Unlike the rest of the theory of falsificationist scientific methodology, the choice of significance level may therefore need to have particular regard to subjective or ethical value choices and the utility of different outcomes[111].

This element of subjectivity runs contrary to Popper's strongly objectivist philosophy of science, but it seems theoretically inescapable in a probabilistic setting. These methodological difficulties are, however, largely alleviated by large volumes of observational data. And Popper naturally had in mind here results from the observation of mass, repeatable physical phenomena. In much, though not all, of natural science, such experimental observation will be available. By the time Popper published *Logic*, he had already witnessed the emergence of very strong experimental evidence in support of probabilistic scientific theories such as the kinetic-molecular theory[112]. But there may be some important branches of natural science, such as the theory of evolution, which have a significant probabilistic element but where controlled observation of mass, repeatable phenomena is not possible.

Falsification of probabilistic hypotheses in the social sciences – where observation may be limited in quantity and non-stationary in quality - may be similarly challenging. This is where these methodological complications will be most acute. These complications have been discussed at some

[110] Popper (1959), p. 195.
[111] This type of problem of decision-making under uncertainty has become a discipline in its own right and is commonly referred to as decision theory. It was pioneered by the American statistician Abraham Wald aound the same time that Popper's Logic was first published - see Wald (1939).
[112] Nye (1972)

length in contemporary scientific methodology literature amidst concern about a ‘replication crisis’ and the growing recognition that research findings based on limited statistical testing may often ultimately prove unreliable[113]. This literature argues that the probability of false positive findings in contemporary scientific research is too high as a consequence of a range of factors: the study design may use a sample size that has inadequate statistical power; modern research may simultaneously test many multiple hypotheses in one study (increasing the likelihood of at least one hypothesis being significant by chance[114]); many research teams may be studying the same phenomenon (increasing the probability that at least one study finds statistically significant data by chance); bias can arise from conflicts of interest or prejudice, and this may be allowed to persist due to equivocal measures of outcomes and / or flexibility in research methodology; the small effect sizes that researchers are attempting to identify in fields such as epidemiology are more prone to false positive conclusions. Some of these factors are fundamentally sociological rather than due to the logic of the methodology. But these factors have more scope to persist in the context of probabilistic hypotheses rather than deterministic ones.

Falsificationism is arguably a highly intuitive approach to developing robust scientific knowledge in a world where the problem of induction has no solution. It recognises that, even in the presence of strong inductive scepticism, a form of (fallible) knowledge can be developed whose strength is based on its relation to empirical observation. But the Duhem-Quine thesis and the application of falsificationist logic in the presence of probability statements serve to highlight that the logic of falsificationism has some important and inherent limitations of a purely logical and philosophical nature (and it will inevitably have further limitations of more practical and sociological natures too).

In the context of our discussion of probability, the other vitally interesting aspect of Popper’s philosophy of science is his assertion that epistemic probabilities cannot be assessed for the truth of a scientific hypothesis. Whilst Popper believed that the degree of corroboration of a hypothesis was non-binary, he did not believe it could be quantified, and he did not equate it with a notion of probability. This philosophical position was and perhaps still remains far from unusual. However, there is another school of thought – broadly speaking, known as the Bayesian school - that argues that it is perfectly natural, logical and useful to measure epistemic probabilities for the truth of a scientific hypothesis, or indeed any other form of synthetic statement. The following section will explore how advocates of an epistemic definition of probability envisioned such probabilities being used in the development of scientific knowledge.

2.2 Bayesianism - an epistemological alternative

“When making a scientific generalisation, we do not assert the generalisation or its consequences with certainty; we assert that they have a high degree of probability on the knowledge available to us at the time, but that this probability may be modified by additional knowledge.”[115] *Harold Jeffreys, Scientific Inference, 1937*.

The first half of this statement is wholly consistent with falsificationist methodology, which explicitly recognises the fallible nature of scientific knowledge. The latter half of the statement, however, marks a divergence in philosophical outlook, at least if the use of the word ‘probability’ above is to be regarded as meaning a measure logically capable of quantitative expression in the context of the truth of a synthetic statement. To the objective falsificationist, the mathematical concept of

[113] See, for example, Ioannidis (2005).
[114] Chapter 5.4, which discusses big data and predictive analytics, will discuss the techniques that statisticians have developed to address multiple hypotheses testing.
[115] Jeffreys (1937), p. 6.

probability is inapplicable to an epistemic statement about a single 'event' such as the truth of a scientific generalisation. However, if we replace the word 'probability' with 'corroboration' and decline to attach a numerical quantity to the degree of corroboration, there would be nothing in Jeffreys' quotation that appears inconsistent with Popper's falsificationism. This section will explore how the epistemic perspective of probability that was discussed in Chapter 1.2 and 1.3 can be applied to scientific inference to obtain probabilities for the truth of scientific hypotheses as per the above quotation of Jeffreys.

As was noted above, the logical approach to probability that was advocated by Keynes and Jeffreys failed to produce general inductive rules capable of generating numerical probabilities for a very broad range of forms of evidence. The sum rule, the product rule and Bayes' Theorem that closely follows them can produce well-defined numerical probabilities for data sets that meet the criteria of exchangeability, assuming a pre-data prior probability distribution can be specified. But beyond that, general rules for the broader enumeration of logical probabilities have proven elusive. Similarly, the logical positivist movement, largely associated with the Vienna Circle of European philosophers of the 1920s and 1930s such as Rudolf Carnap also set out to show how partial evidence could be rationally treated to obtain objective positive measures of probable knowledge. This program never achieved its authors' ambitious objectives, thought its attempts and writings had a very significant impact on philosophy of science and an influence beyond the natural sciences. Consequently, in the 20th century history of the philosophy of science, the objective of developing a positive theory of how to use inductive evidence in the testing of scientific theory saw the logical approach of Jeffreys and Keynes largely superseded by the subjective probability philosophy of Ramsay and De Finetti.

In more recent decades, however, the logical approach has made something of a resurgence. When we refer to Bayesianism in this section, we therefore have mainly the logical probability in mind, though similar points will often apply to the subjectivist approach – the essential point is that probability here is being used as an epistemic measure of plausibility for singular statements rather than only as a frequency or propensity for the occurrence of repeatable mass phenomena. The passage below from Jeffreys' *Theory of Probability* highlights his philosophical dispute with those who reject the validity of epistemic probability and its application to scientific hypotheses:

"The most serious drawback of these [frequency] definitions [of probability] is the deliberate omission to give any meaning to the probability of a hypothesis. All that they can do is to set up a hypothesis and give arbitrary rules for rejecting it in certain circumstances. They do not say what hypothesis should replace it in the event of rejection, and there is no proof that the rules are the best in any sense. The scientific law is thus (apparently) made useless for purposes of inference. It is merely something set up like a coconut to stand until it is hit; an inference from it means nothing, because these treatments do not assert that there is any reason to suppose the law to be true, and it thus becomes indistinguishable from a guess."[116]

This passage does not make explicit reference to Popper or falsificationism, but there can be little doubt that this is what he has in his sights here – to Jeffreys, it is falsificationism that sets up a hypothesis 'like a coconut to stand until it is hit'. The key point that Jeffreys' quotation above makes is that falsificationism does not consider the performance of a hypothesis *relative* to another – if the hypothesis is falsified, the scientist is potentially left with nothing until the day a new, better theory comes along, which might never arrive. The Bayesian approach, by assuming the probability of the truth of a hypothesis can be quantified, naturally permits such probabilities to be assessed for all possible hypotheses under a given set of evidence. This, in principle, can provide an assessment and

[116] Jeffreys (1961), p. 377.

a ranking and comparison of all these hypotheses at any given moment. This is arguably a more informative approach to appraisal of hypotheses than working from the assumption that a given hypothesis is the 'right' one (and therefore all others are the 'wrong' one) until it reaches the point of falsification.

Philosophical Scepticism on the Measurement of Epistemic Probabilities for Scientific Hypotheses

Finding an approach to logically quantifying a probability for the truth of a hypothesis is, however, a seriously demanding problem. It requires that all forms of evidence relevant to the probability can be quantified; and that a functional form can be found that can transform these quantitative measures into an objective epistemic probability. Many philosophers of science have expressed severe doubts that such a thing can be done, even amongst those who have accepted and embraced the general development of epistemic, logical probability.

As noted in Chapter 1, Keynes, whilst a forthright advocate of probability as a measure of partial knowledge, did not believe that all such probabilities could be enumerated and thus compared and ranked. And many leading 20th century philosophers of science, including those sympathetic to idea of epistemic probability, rejected the idea that epistemic probability could be usefully applied to the truth of a scientific hypothesis. For example:

- Ernest Nagel, another notable philosopher of science who was a general proponent of the use of epistemic probability, was highly sceptical that probabilities for scientific hypotheses could be enumerated and ranked[117].
- Rudolf Carnap was equal only to Jeffreys in his commitment to the development of an epistemic probability that was based on inductive inference. But even he explicitly viewed it as 'out of the question' that quantitative epistemic probabilities could be attached to the degree of confirmation attributable to complex scientific theories such as quantum theory or Einstein's general relativity. In his view, the relevant body of evidence was so immense and complex that it would be impossible to express it in a quantitative form that was capable of use in a piece of inductive logic that would produce an objective epistemic probability[118].
- The Cambridge philosopher R.B. Braithwaite who we discussed earlier was another notable philosopher of science of the mid-twentieth century who expressed strong scepticism that a theory of logical probability could enumerate probabilities for hypotheses. Braithwaite wrote 'it is difficult to see how a formal logic of credibility or acceptability or confirmation can help in the matter, beyond pointing out obvious truisms as that if a hypothesis is supported by good evidence, any lower-level hypothesis which is a consequence of it is also supported by good evidence and may well also be supported by other good evidence.'[119]

Producing Epistemic Probabilities for Scientific Hypotheses: The Basic Approach

So, how did Jeffreys and the wider Bayesian school approach the evaluation of the probability of a scientific hypothesis? Well, the philosophical starting point is to *assume* it is possible: as noted in Chapter 1, Jeffreys' probability theory took it as axiomatic that for all propositions and any given set of evidence, there exists a unique conditional probability.

From there, the basic mathematical starting point is Bayes' Theorem, which, as noted in Chapter 1, is sometimes referred to as the *Principle of Inverse Probability* in this context. For our current purposes, it can be conveniently written in the form:

---

[117] Nagel (1939)
[118] Carnap (1950), p. 243.
[119] Braithwaite (1953), p. 197.

$$P(h_r|e) = \frac{P(e|h_r)P(h_r)}{\sum_{i=1}^{n} P(e|h_i)P(h_i)}$$

where h_1, h_2,…, h_n are mutually exclusive and exhaustive hypotheses (that is, one and only one of the hypotheses is true); and *e* is the available evidence. In standard Bayesian terminology, P(h) is the prior probability of the hypothesis *h* being true (i.e. prior to the specific evidence *e* being available, but after consideration of all other available evidence and relevant background knowledge and information); and P(h|e) is the posterior probability after consideration of evidence *e*.

In the vernacular of falsification, evidence that falsifies one of the competing hypotheses would reduce its posterior probability to zero and thereby reduce the field of surviving possible hypotheses from *n* to *n-1*. Clearly, if this falsification process occurred for all hypotheses until only one possible hypothesis remained, then the posterior probability of that hypothesis would increase to 1. But, more generally, it is not necessarily the case that evidence that falsifies one of the *n* hypotheses will necessarily result in an increase in the posterior probability of *all* the remaining hypotheses (though clearly the posterior probabilities of the surviving hypotheses must sum to one, if it is assumed that our set of hypotheses is the exhaustive set of all possible hypotheses).

The assumption that it is known that there are exactly *n* and only *n* hypotheses with a non-zero probability of being true may seem innocuous. But it may only be reasonable under fairly special circumstances. It would seem especially difficult to justify this assumption when considering the behaviour of relatively poorly understood phenomena. How can it be known *with certainty* that the truth resides amongst the candidate hypotheses? How can we be certain that there isn't some other hypothesis that we have completely overlooked that turns out to be true? And doesn't this imply that all future potential scientific progress must be anticipated? For example, suppose we are in 17th century Cambridge, busying ourselves with a Bayesian appraisal of various competing theories of gravity (let us set aside that Bayes has not yet been born!). According to this framework, some non-zero probability ought to be attached to the theory of relativity. There would be no particular evidential support for it at the time, so its probability may not be very high, but it must be non-zero, otherwise the process of updating the posterior probabilities could never attach any probability to the theory of relativity, no matter what evidence subsequently arose to support it.

Priors and Simplicity

In many cases, it may be natural to assume the number of candidate scientific hypotheses *n* is infinite – after all, there may be an infinite number of hypotheses that fit a specified body of scientific evidence (to draw a simply analogy, there are an infinite number of curves that fit through any finite series of points). The probabilities of the possible hypotheses, however, must of course sum to 1. This is true for both the posterior and prior probabilities. Bayes' Theorem will ensure that the posterior probabilities sum to 1 if the prior probabilities do. How do we ensure the prior probabilities sum to 1 when an infinite number of hypotheses have a non-zero prior probability?

Jeffreys proposed a novel general solution to this problem. He argued that the hypotheses' prior probabilities could be ordered (as, by his own axiom, all probabilities can be ordered), and they must form a converging series that, naturally, summed to 1. He noted this implied the order of magnitude of the prior probabilities must diminish quite rapidly. The novelty of his approach was to then propose that the prior probabilities of the hypotheses could be ordered by the *simplicity* of the hypothesis. According to Jeffreys, this ordering could be achieved because all scientific hypotheses could be represented as 'differential equations of finite order and degree, in which numerical coefficients are integers'. Complexity (the antithesis of simplicity) was defined as 'the sum of the order, the degree, and the absolute value of the coefficients'.[120]. He then took it as axiomatic that a simpler hypothesis had a much higher prior probability than a complex hypothesis, such that the sum of the infinite series of increasingly complex hypotheses would converge to one. Jeffreys called this axiom the *Simplicity Postulate*. This postulate also provides a solution to the difficulty we noted above of having to anticipate all future scientific progress: if all possible hypotheses can be expressed in a functional form as Jeffreys describes, then it is indeed possible for all future potential theories to be anticipated and for a prior probability to be attached to all of them.

The Simplicity Postulate is not logically derived from more basic premises. It is postulated as an axiomatic assumption that Jeffreys believed was reasonable given the scientist's strong preference for simple hypotheses. And, perhaps predictably, other philosophers have argued that it is not possible to quantify and rank the simplicity of all forms of scientific hypothesis, either in the way described by Jeffreys or indeed any other way[121].

It can be seen from the above brief discussion that the foundations of the Bayesian framework rely on a number of arguably quite demanding axioms and assumptions. The ability to make an exhaustive specification of all possible hypotheses; the existence of a unique numerical probability for any hypothesis given any form of evidence; and, related to this, a means of expressing ignorance through the specification of uninformed priors. These are all philosophically stretching and contested axioms for a system of probability as a degree of partial belief in a scientific hypothesis.

Whether invoking the Simplicity Postulate or not, the Bayesian framework and the Principle of Inverse Probability means that the prior probabilities of hypotheses must inevitably enter into their posterior probabilities. Jeffreys recognised this was generally unattractive, and he argued for a scientific method that minimised the dependency of the posterior on the prior:

"Prior probabilities enter into our formulae, but we do not know their values, and they always affect the posterior probabilities. But in scientific work, though we can never make the posterior probability completely determinate, we can make it so near zero or unity as to amount to practical certainty or impossibility for *all* ordinary values of the prior probability. This is done by repeated verifications and crucial tests."[122]

Philosophical Differences and Practical Similarities

It is interesting to note how similar Jeffreys' final sentence above is to a falsificationist view of empirical science. Clearly, a 'crucial test' is an experiment or observation that could rule out the hypothesis if a particular result was obtained – that is, which falsified the hypothesis. Jeffreys repeatedly emphasised the importance of such tests:

"The more facts are in agreement with the inferences from a law, the higher the probability of the law becomes; but a single fact not in agreement may reduce a law, previously practically certain, to the status of an impossible one."[123]

In Bayesian terminology, 'a single fact not in agreement' implies $P(e|h) = 0$. By the Principle of Inverse Probability, any such result renders the posterior distribution of the hypothesis (i.e. $P(h|e)$) zero also. Falsificationists call this a falsification. Both approaches arrive at the same conclusion that

[120] Jeffreys (1937), p. 45-46.
[121] Hacking (1965), p. 225.
[122] Jeffreys (1937), p. 23.
[123] Jeffreys (1937), p. 9.

the probability of the hypothesis being true is zero (subject to the Duhem-Quine thesis caveat discussed above which again applies similarly to both philosophical perspectives).

Given the vigour with which the objective versus Bayesian probability debate has historically been conducted, it is perhaps easy to overlook that, if we set aside the philosophical dispute about whether or not we should attach numerical probabilities to scientific hypotheses, we can find some significant similarities between the advocated scientific methods of the falsificationist and Bayesian approaches. Both approaches advocate a vigorous approach to empirical testing of a scientific hypothesis that values both the quality and quantity of observations. The inductive logic of the Bayesian setting means that each positive observation increases the posterior probability of the hypothesis. Popper also acknowledged (see Chapter 2.1 above) that increasing numbers of positive observations increased the degree of corroboration of a hypothesis. In the Bayesian approach, by quality here we mean an observation that will have a large impact on the ranking and differentiation amongst the posterior probabilities of the various competing hypotheses. In general, this occurs when the differences between $P(e|h)$ and $P(e)$ are significant. In the falsificationist approach, quality refers to the testing of the hypothesis's bold predictions that can readily expose it to falsification (i.e. where $P(e|h)$ is zero for some *h*). These are not inconsistent views on the relevance of different types of evidence. In short, *these distinct philosophical perspectives value the same forms of evidence in their approaches to the empirical testing of hypotheses.*

Finally, even an ardent Bayesian scientist and philosopher such as E.T. Jaynes conceded that numerical quantification of the probability of a hypothesis is, in real life, usually beyond us:

"In practice, the situation faced by the scientist is so complicated that there is little hope of applying Bayes' Theorem to give quantitative results about the relative status of theories."[124]

In this light, a relative preference for the use of either degrees of corroboration or the posterior probability of a scientific hypothesis appears to be of little obvious consequence for how evidence is used to determine the degree of confidence in a hypothesis.

Similarly, both philosophical outlooks *prefer simplicity in a scientific hypothesis*. In Bayesian terms, this preference may be expressed by the postulate that a simple hypothesis will have a higher prior probability than a complex hypothesis. In falsificationism, simple hypotheses are preferred to complex ones because they are argued to be more falsifiable (as they rule out more) and are hence more corrobarable.

The above observations can be put another. The logic of Bayesian updating is not disputed, and it is quite possible for scientists or methodologists to naturally use Bayesian logic *qualitatively* without subscribing to the more ambitious view that the prior and posterior probabilities can be quantified. As an example, we can consider the historical curiosity from the 1920s of Henry Ludwell Moore's Venus theory of the business cycle[125]. This theory proposed that the orbital patterns of the Earth and Venus caused a weather cycle which in turn caused the business cycle. It, unsurprisingly, struck many of his peers as somewhat fanciful, and a Bayesian logic was implicitly employed in the rejection of the relevance of whatever evidence Moore presented:

---

[124] Jaynes (2004)

[125] See Morgan (1992), pp. 26-33 for a fuller discussion of Moore's theory and the statistical evidence he gathered to support it.

"Even if a statistical test should yield a very high correlation, the odds thus established in favour of such a hypothesis would have to be heavily discounted on account of its strong *a priori* improbability."[126]

This is an example of how, when scientists have background information that is not directly related to the evidential data, the Bayesian approach provides an explicit mechanism – the prior probability - for incorporating this information into the assessment of the hypothesis. There is no parallel explicit mechanism for incorporating this qualitative background information into falsificationist methodology. Nonetheless, it would be difficult to maintain the view that falsificationist methodology would fail to establish that the Venus theory is a weak one (for example, it is not part of a deductive system that offers direct or indirect empirical corroboration).

The Bayesian approach is, in an epistemological sense, more ambitious than falsificationism – it promises more. As noted above, one of its promises is that it provides a means, at least in theory, of *ranking competing hypotheses*. A limitation of Popperian falsificationism is that there could be many competing hypotheses that have not (yet) been falsified by the available evidence. In such circumstances, which hypothesis should be used? The falsificationist would argue that the theory that has been corroborated to the higher degree should be the preferred one. But as degrees of corroboration are not always numerically orderable, this may not imply a unique ranking. Of course, in such a case a falsificationist may argue that the rankings of posterior probability obtained by the Bayesian calculations are arbitrary and unreliable. And from the above reading, it seems likely that both Popper and Jeffreys would agree that it is necessary to conduct further well-designed experiments that are capable of further differentiating between (and ideally falsifying some of) the competing hypotheses.

Statistical Testing Again

A parallel to the above falsificationism-Bayesianism debate around the testing of scientific hypotheses arises in the broader context of statistical inference. We saw above that Popper advocated the use of a significance test as a means of 'practical falsification' in the case where the deductions of the scientific hypothesis were probability statements (as such statements cannot be logically falsified). We noted that this amounted to a statistical significant test or hypothesis test, which is a well-established technique of statistical inference that was first developed by Sir Ronald Fisher in the early decades of the twentieth century.

Significance testing, as developed by Fisher and Pearson in the first half of the twentieth century, can be viewed as a form of practical falsificationism that has much wider application than the testing of scientific hypotheses. In significance testing, as in the falsificationist scientific method, no probability of 'truth' is ever estimated for the null hypothesis, but the hypothesis is rejected as false if the evidence against it is sufficiently strong (that is, if the probability of the evidence arising given the hypothesis is true is sufficiently small).

The fallibility of significance testing is fundamental and unavoidable. This was explicitly acknowledged by Fisher in his advocacy of the approach. As he put it when discussing the logic of significance testing: "Either an exceptionally rare chance has occurred, or the [null hypothesis] is not true"[127]. He also wrote:

"If we use the term rejection for our attitude to such a hypothesis [one that has been rejected in a significance test], it should be clearly understood that no irreversible decision has been taken; that,

[126] Hotelling (1927), p. 289.
[127] Fisher (1956), p. 39.

as rational beings, we are prepared to be convinced by future evidence that appearances were deceptive, and that in fact a very remarkable and exceptional coincidence had taken place."[128]

The above remarks, made in the context of statistical hypothesis testing in general, are very similar to those that arose in the discussion of the 'practical falsification' of probabilistic scientific hypotheses. There it was noted that the test is fallible and that a rejection or practical falsification must be provisional as it was always theoretically possible for further evidence to arise that logically demands the cancellation of the rejection.

The significance testing approach to statistical inference was founded on a scepticism about the ability to obtain rational (non-arbitrary) estimates for the probabilities of a statistical hypothesis being true. Quoting Fisher a final time: "While, as Bayes perceived, the concept of mathematical probability affords a means, in some cases, of expressing inferences from observational data, involving a degree of uncertainty, and of expressing them rigorously, in that the nature and degree of the uncertainty is specified with exactitude, yet it is by no means axiomatic that the appropriate inferences, though in all cases involving uncertainty, should always be rigorously expressible in terms of the same concept"[129].

Bayesians could retort that the Fisherian approach inevitably involves at least as much arbitrariness as the Bayesian approach that Fisher objects to in the above quotation. In particular, the choice of null hypothesis, the choice of test statistic and the choice of significance level are all, to some degree, arbitrary, and all can alter the conclusion of a significance test[130].

Ultimately, conjecturing from a finite set of observations to a universal generalisation or other inference about a population must involve some assumptions, some form of judgement, and perhaps some arbitrary choices. Beyond the philosophical debates, it is clear that vigorous empirical testing, simplicity in hypotheses, and a recognition of the fallibility of scientific knowledge are all virtues of the scientific method that both objectivists and Bayesians provide a rationale for (and can each claim as their own!).

2.3 Causation and scientific explanation

So far, our discussion of science and its methods suggests that the aim of science is to reliably and objectively identify and accurately quantify recurring relationships between observable phenomena, and to describe those relationships as economically and simply as possible. The recurring nature of these relationships allows us to infer from the observed to the unobserved. Such inferences will not be infallible, but can create a form of knowledge that is capable of very reliable predictions within understood domains. But, as powerful as this descriptive and predictive knowledge may be, this description of scientific knowledge seems to be missing an important dimension: isn't science about explaining *why* as well as describing *what*?

This leads us to the vexing subject of scientific explanation and causation: what does it mean to *explain* a relationship between two phenomena? To identify the cause or causes of a type of occurrence? Can it be done and, if so, how? Ernest Nagel provided a seemingly a simple definition. To identify something's causes means 'ascertaining the necessary and sufficient conditions for the

[128] Fisher (1956), p. 35.
[129] Fisher (1956), p. 37.
[130] Howson and Urbach (1993), Chapters 8 and 9 for a Bayesian critique of frequentist statistical inference. This book, more generally, provides a comprehensive exposition of philosophical underpinnings of the Bayesian approach to statistical inference.

occurrence of phenomena'[131]. The following discussion explores this idea further, and highlights some of the philosophical debate that surrounds this topic.

Following Nagel's definition, we may identify C as a necessary and sufficient cause of some effect E if the occurrence of C is *always* followed (in time[132]) by the effect E, whilst the absence of C, all other things being equal, which we will refer to as not-C, is *always* followed by the absence of E, which we will denote here not-E. In passing, we may note that if E always occurred when C occurred, and E sometimes occurred when C did not occur, this is known as 'causal overdetermination', and means one of a number of different causes may be sufficient to cause E. In this case C is a sufficient cause but is not necessary.

The values and states of the other conditions and circumstances referred to as 'all other things' in the previous sentence can be denoted by K for ease of exposition. K is sometimes referred to by philosophers as the *causal field*. It essentially means all causally relevant background. K can include factors that are naturally unchanging (e.g. the gravitational constant) or those that are assumed to by unchanging within a given theory (e.g. the money supply) or those that are assumed to be unchanging in a very particular way, that is, to be absent (e.g. friction). C may be followed by E (and not-C followed by not-E) under all possible conditions of K, or only for some sub-set of possible values of K. The above definition of the necessary and sufficient cause, C, has assumed that the relationship applies for all possible values of K. If the relationship holds for only some K, then C is not sufficient, and may or not be necessary, for the occurrence of E.

This merely *describes* a relationship between two phenomena, C and E, and the conditions K under which the relationship does and does not hold. To be recognised as a *causal* relationship or scientific explanation, we might also expect some explanation of *why* E is always followed by C, and *why* not-C is always followed by not-E. For example, it has been known for many centuries that the ebb and flow of the tides follows a pattern related to the lunar cycle. But there was no good answer to the 'why' question until Newton's theory of gravitational attraction[133].

Correlation is not causation…so what is causation?

What does a good scientific explanation look like? As we shall see below, one answer is that a scientific explanation is of the form of a logical deduction that shows that E *must necessarily* follow C, given some stated premises. In the absence of such an explanation, we merely have a description of a correlation. And is all statistics students know, correlation is not causation.

The distinction between a causal explanation and a 'mere' empirical generalisation has been the subject of philosophical discourse for centuries, and especially since Hume[134]. In the early twentieth century, the Cambridge philosopher W.E. Johnson referred to these two forms of statement as 'universals of law' (a causal explanation in the form of a logically deduced law-like statement) and 'universals of fact' (empirically-observed recurring relationships)[135]. But according to Hume, there really is no distinction between these two forms of statement – that is not to say that effects do not

---

[131] Nagel (1959), p. 382.

[132] Philosophers can seem quite particular about the notion of time when discussing cause and effect. Naturally, it is generally (though not universally) accepted that an effect cannot occur before a cause. A philosopher may also stipulate that there is no time-interval between the cause event and the effect event, as the 'action at an interval of time' would essentially be inexplicable. So, cause and effect may then be characterised as a 'temporally continuous regular sequence' or chain of events. See Braithwaite (1953), p. 309.

[133] This example was given in Salmon (1984).

[134] Hume (1740), Hume (1748)

[135] Johnson (1924)

have a cause, but rather that the fundamental cause is unobservable and essentially unknowable. Our knowledge is only capable of determining the empirical observation of a 'constant conjunction' of two phenomena. According to Humean epistemology, the notion of causal explanation is merely psychological rather than epistemological. Deductive scientific theories and laws of nature do not depict the truth; they merely provide succinct and efficient ways of *describing* the empirical regularities that we observe, nothing more. That is, 'universals of law' are just efficiently statements of 'universals of fact'. This is the nature and limit of scientific knowledge, according to Hume.

Consider again the Newtonian explanation of the tides given above from this perspective. A Humean rebuttal of the Newtonian explanation might argue that Newton's laws identified and quantified the phenomenon of gravitational attraction, but they did not explain *why* this phenomenon exists. A law-based explanation is required for the explanatory law, and so we reach an infinite regress. The distinction between an empirical generalisation and scientific explanation, it may be argued, is therefore a difference of degree rather than kind – explanation becomes hierarchical, like Braithwaite's model of scientific hypotheses. And the Humean philosopher would argue that the ideal of causal explanation can never be obtained.

A school of philosophers has maintained Hume's empirical scepticism in modern times. Indeed, it was probably the dominant philosophical outlook on scientific explanation until the middle of the twentieth century and the influential work of the logical empiricists. For example, writing in the early 20th century, Bertrand Russell famously dismissed the philosophical notion of causality as 'a relic of a bygone age, surviving like a monarchy, only because it is erroneously supposed to do no harm'[136]. But modern philosophers have increasingly found this position to be inadequate. In a nutshell, Hume's empirical scepticism sets the linguistic bar for 'causation' and 'explanation' at a very high level – it is used to mean certain knowledge or proof of the reason why; and a less severe definition may usefully serve to recognise important forms of imperfect explanation.

The Deductive-Nomological Model of Scientific Explanation

Hempel and Oppenheim in 1948, in a highly influential paper that re-ignited philosophical interest in the topic of scientific explanation and causation, argued that a legitimate scientific explanation must feature a logical deduction from a general or universal law that has empirical content[137]. This became known as the deductive-nomological model of scientific explanation. A quite similar position was advocated a few years later by the Austrian logical empiricist Herbert Feigl, who also defined causation as something delivered by a scientific law: "Causation is defined in terms of predictability according to a law (or, more adequately, according to a set of laws)."[138]

So, from this logical empiricist perspective, the general structure of the scientific explanation of a specific event will require as premises a set of initial conditions or circumstances together with some form of (empirically corroborated) law(s) and / or generalisation(s) from which the occurrence of the event, E, can be deduced. Feigl, Hempel and Oppenheim viewed explanation and prediction as logically equivalent and only different in terms of the direction of time in which they worked – that is, explanation (of past observations of some phenomena) is prediction (of the future behaviour of the phenomena) in reverse, and vice versa. This is sometimes referred to as the symmetry thesis. It is a point that is arguably more germane to the social rather than natural sciences, and we will therefore return to this point in Chapter 3.

---

[136] Russell (1918), p.180.
[137] Hempel and Oppenheim (1948)
[138] Feigl (1953), p. 408.

From as early as Aristotle, some philosophers have argued that the premises of an explanation must be 'known to be true'[139]. As has been discussed above, the laws produced by scientific theories can be used to make deductive predictions about the behaviour of phenomena. Virtually all philosophers of science would agree with that. And whilst scientific laws are developed by deductive logic, these laws are *not* known to be certainly true as their premises may not be correspond precisely with the 'real-world'. So the epistemic quality of scientific explanation appears to inevitably occupy a sort of halfway house between, on the one hand, the 'pure' deductive knowledge that arises from premises known to be true in the sense of pure logic or mathematical truths, and, on the other hand, the mere inductive enumeration of positive instances (which Hume's problem of induction argued was, on its own, epistemologically worthless).

Causation and a Uniformity of Nature Postulate

If we do accept the use of empirically-corroborated deductive theories as scientific explanations, it is important to note that, when used for prediction of future phenomena, it is necessary to invoke some form of principle of causality. Taking a famous example of this form of assumption, John Stuart Mill invoked a principle of the uniformity of nature as follows:

"There are such things in nature as parallel cases; that which happens once, will, under a sufficient degree of similarity of circumstances, happen again."[140]

Arguably, Mill's principle merely provides a circular definition for what constitutes 'a sufficient degree of similarity of circumstances' – that is, they are the circumstances in which something that has happened once will happen again. Nonetheless, its essence is clear: scientific knowledge relies on some form of uniformity of nature in time and space, and, given such uniformity, it is reasonable to assume the causes that led to given effects in the past will do so again in the future.

This uniformity of nature postulate was further expanded upon by John Venn towards the end of the nineteenth century, when he described it as the assertion that: 'wherever any two or more attributes are repeatedly found to be connected together, closely or remotely, in time or in space, there we have a uniformity'[141].

Again, we might question how we *know* that these attributes are always bound together. This is really Hume's problem of induction again. On this Venn was clear, the problem cannot be solved, only assumed away:

"I am very decidedly of the opinion that the difficulty does not admit of any *logical* solution. It must be assumed as a postulate, so far as Logic is concerned, that the belief in the Uniformity of Nature exists, and the problem of accounting for it must be relegated to psychology."[142]

The uniformity of nature postulate makes a form of inductive inference from observed to unobserved possible. It enables the link between causation, explanation and prediction. If we take this position, then well-tested and accepted scientific theories will generally be accepted as offering scientific explanations of the behaviour of related phenomena. Again, this is fundamentally consistent with both the falsificationist and Bayesian philosophies of scientific method discussed in Chapters 2.1 and 2.2 – indeed, it is the essence of what science does.

---

[139] Nagel (1979)
[140] Mill (1879), Book 3, Chapter 3, Section 1.
[141] Venn (1889), p. 93.
[142] Venn (1889), p. 132.

With the acceptance of this broader definition of scientific explanation, there are a couple of (related) key dimensions in which the quality of an explanation may be assessed. One dimension is the degree of empirical confirmation for the causal laws that are used in the explanation. Falsificationists and Bayesians may differ in their perspective on how to describe empirical confirmation, but both schools would generally argue that it is at least sometimes possible to order the degree of confirmation of a theory, and that the empirical testing of the theory is fundamental to the scientific method.

The second important dimension in the quality of a scientific explanation is the generality of the law or system of laws from which the explanation is deduced. As discussed in Chapter 2.1, a deductive system can be thought of as a hierarchy, with the most general (and theoretical) hypotheses at the top, potentially leading to many varied lower level hypotheses that can each be directly empirically tested. The more general and higher level the system of hypotheses is that delivers the deductive explanation, the greater the logical strength of the deductive system and the more satisfactory the explanation. This notion of logical strength is related to the desideratum of simplicity that is common across virtually all forms of philosophy of science – a more powerful explanation is one that can explain more with less. Furthermore, these two dimensions of the quality of a scientific explanation are related – where higher level hypotheses engender many lower level hypotheses, the empirical confirmation of each lower level hypothesis can be viewed as indirect evidence in support of the others.

The above discussion of causation has only considered the simplest of settings: in particular, it has, at least implicitly, suggested that every effect *has* a cause and so the difficulty merely lies in knowing what that cause is (causal determinism); and also we have thus far assumed that causation is *universal* - that is, the same causal circumstance(s) *always* lead to the same effect. We now explore the philosophical consequences of generalising beyond these two simplifying (and related) assumptions.

Causation and Determinism

Determinism is the philosophical idea that complete knowledge of a physical system will allow all the behaviour of all phenomena therein to be uniquely determined for all time (past, present and future). Determinism implies all effects have a cause that it is at least theoretically knowable.

Classical mechanics was developed in full accordance with a deterministic view of the physical world. For example, given knowledge of the current mechanical state of a system (the initial mass, velocity and position of the objects of a system together with the forces acting upon that system), Newton's laws of motion uniquely determine the mechanical state of the system at any future time. It was the explanatory power and extraordinary predictive reliability of classical mechanics that inspired Laplace's enthusiastic and famous embrace of determinism:

"An intelligence knowing all the forces acting in nature at a given instant, as well as the momentary positions of all things in the universe, would be able to comprehend in one single formula the motions of the largest bodies as well as of the lightest atoms in the world."[143]

In this setting, uncertainty or merely probabilistic relations are only expressions of human ignorance. Perfect knowledge implies perfect foresight (and perfect understanding of the past). The precise determinism of Newton and Laplace's classical mechanics was somewhat muddied by the profound new developments that emerged from late 19th century physics. In particular, the development and application of statistical mechanics in highly successful theories across fields such as

[143] Laplace (1820), Preface.

thermodynamics and the behaviour of gases represented an important departure from the classical methods. Statistical mechanics, like classical mechanics, was based on Newton's laws of motions. However, statistical mechanics applied them to aggregate numbers of molecules instead of individual objects. The mechanical states of huge number of molecules could not be individually specified, and so the initial state of the system was instead specified by assuming a probability distribution for the states of the molecules and a stochastic process for how these states changed over time (such as Brownian motion).

Whilst statistical mechanics made explicit use of probabilistic assumptions, its foundation was still the deterministic laws of classical mechanics. Probabilities entered into the analysis simply because it was not practical to specify individual states at the molecular level. The introduction of quantum theory in the early 20th century, and its impressive empirical corroboration across a wide range of scientific fields, produced a more fundamental challenge to the notion of mechanical determinism and causation. Whilst probability entered statistical mechanics as a matter of practicality, in quantum theory probability played an essential and unavoidable theoretical role. According to this theory, some phenomena of the physical world (and therefore the laws that describe them and their relations with each other) are *irreducibly statistical*. Complete knowledge of the physical system could therefore no longer deliver perfect foresight or a deterministic description of all phenomena at any moment in time. Quantum theory describes a world where not all effects have a knowable cause.

Probabilistic Causation

In our discussion so far, we stated that C was a cause of E if, in some given a set of known conditions, K, C is always followed by E, and not-C is always followed by not-E. We now consider the case where C makes the occurrence of E merely *more probable*. Let's refer to this idea as *probabilistic causation*. Perhaps surprisingly given the extensive use of probability in scientific theories since the nineteenth century, the philosophical study of probabilistic causation didn't really get underway until 1962 when Hempel attempted to extend his (deterministic) Deductive-Nomological (D-N) model into a probabilistic setting[144].

Philosophers generally consider that, like deterministic causation, probabilistic causation also requires an empirically-corroborated scientific law to be used in a scientific explanation. However, in this case, the scientific law is a *statistical law rather than a universal law* (for example, the kinetic theory of gases deduces a probability distribution[145] for the velocity of particles in idealised gases). Therefore, in some instances of C, C may be followed by not-E, and not-C may be followed by E. But, given C, the occurrence of E is *more probable* than it is given not-C. That is, $P(E|C) > P(E \mid \text{not-}C)$.

In this case, C is neither a necessary nor sufficient condition for E. Note that, under this definition, the conditional probability of E occurring given C could be very low in absolute terms. Even in such cases, any difference between conditional and unconditional probabilities would be enough to deem C to be causally relevant. But this is a matter of philosophical taste. When Hempel developed probabilistic causation using statistical rather than universal laws in 1962 (which he called the inductive-statistical (I-S) model of explanation), he insisted a 'high' probability must be generated for the effect, E, for C to be considered a cause. In contemporary philosophy of science, the requirement for a 'high' probability (rather than merely an increase in the probability resulting from the presence of C), is generally regarded as unnecessarily restrictive (and, in later years, Hempel

---

[144] Hempel (1962)

[145] The Maxwell-Boltzmann distribution, which is a chi distribution with three degrees of freedom.

himself also converted to the view that probabilistic causation merely requires an increase in the conditional probability, without reference to its absolute value)[146].

Note that the specification of K may have a crucial impact on how the conditional probability of E changes in the presence of C – the impact of C on the conditional probability of E may even reverse direction under different values for K. Probabilistic causation can therefore become quite counter-intuitive if all the relevant conditions K are not identified and held constant. There are some famous examples. For instance[147], birth control pills are known to contain blood clotting chemicals that increase the risk of thrombosis. And yet empirical studies show that those who take the pill have a lower probability of developing thrombosis than those who do not (relative to a control group of women of similar age, etc.). The reason for this is that pregnancy increases the probability of thrombosis. If K includes the state of pregnancy (i.e. either all are pregnant; or none are pregnant), then the result will show that the pill increases the probability of thrombosis. But if this factor is not held constant, the opposite result will arise, because the birth control pill-taking population has such a lower proportion of pregnant people than the control group.

The above case where probabilistic relationships are deduced from a scientific theory can be contrasted with another case where there is no theoretical causal argument but merely an (as yet) unexplained empirical statistical generalisation. That is, it might be possible to establish an empirical probabilistic relationship between C and E, and between not-C and not-E, without having a deductive theory that explains why it occurs. Most philosophers of science would argue this does not constitute a scientific explanation.

This type of situation – the observation of some empirical association without a fundamental theoretical explanation – is especially common in the social sciences and we will discuss this point further in that particular context in Chapter 3. However, the establishment of these empirical relationships can often be a crucial point in the development of scientific theories in the natural sciences too – these are the empirical observations that can motivate the development of a theory. A well-known example would be Boyle's Law, which states that the pressure of a given mass of gas is inversely proportional to its volume at a fixed temperature. This law is a statement of an empirical relationship, it is a description, not a scientific theory that offers an explanation for *why*. The relationship it describes has since been logically deduced in the kinetic theory of gases, thus providing the scientific explanation.

This distinction between an empirical generalisation or experimental law and a scientific theory, naturally prompts the question – is there a clear distinction between these two concepts? The essential difference between them is that the empirical generalisation is based on entirely observable and empirically determinable phenomena. A scientific theory, on the other hand, will entail a level of abstraction, typically in the form of variables and phenomena that are 'theoretical' or at least not directly observable.

Whilst the essence of this distinction would seem intuitive, it may be less sharp in real-life. It is quite difficult to measure an empirical generalisation in a perfect theory vacuum. The example of Boyle's Law above requires clarity about what we are measuring – what is pressure, temperature and volume? The nature and behaviour of these phenomena may themselves be the subject of scientific theories. Nonetheless, the distinction between the measurement of empirical relationships and the explanation (or, if a sceptical empiricist, a more efficient description) of those relationships through

[146] Salmon (1984), p. 190.
[147] Hesslow (1976). See Hoover (2001), Chapter 1.2 for further discussion.

fundamental theoretical ideas is an important one when considering the different stages of the scientific method, and, in particular, how they differ across the natural and social sciences.

The epistemological 'strengths' of an empirical generalisation and a scientific theory also differ. The empirical generalisation may be taken as a statement of reality that is unchanging (at least in the stationary environment which is generally found in the natural sciences, perhaps less so in the social sciences). The scientific theory that explains or describes the generalisation, in contrast, is always considered fallible, and subject to revision and improvement.

Probabilistic Causation and Indeterminism

Recall from our discussion of deterministic causation above that the state not-C was defined as one where all other factors and circumstances, K, are the same as in state C. This creates a potential philosophical difficulty for probabilistic causation. If *all* causally-relevant conditions are *exactly* the same, what makes the relationship between C and E probabilistic rather than deterministic? One interpretation of probabilistic causation is that some of the causal field, K, is varying in an uncontrolled way. These variations must be random, otherwise they would be identified in the causal structure.

If we think of cause as a disposition or propensity, then the specification of K in probabilistic causation becomes analogous to the reference class in objective probability, and this philosophical difficulty is similar to the reference class problem discussed in Chapter 1 - holding the causal field entirely fixed is analogous to a reference class of 1, and implies determinism. And, as in the reference class problem, an *epistemic* form of solution may be offered. That is, the reason the causation is only probabilistic rather than deterministic, is because we do not fully *know* what variables are relevant to the causal field, K, and what values they are taking from one instance of C to the next. This implies that the probabilistic version of causation and explanation is an incomplete version of the deductive-nomological explanation: that is, with complete knowledge of the causal field, K, the causal relation between C and E is a deterministic one.

But the epistemic argument that probabilistic causation is always a merely incomplete version of deterministic causation assumes a deterministic world. In an indeterministic world where some relationships between physical phenomena are irreducibly statistical, it would not be possible to ever know the values of the entire causal field at any moment in time (quantum theory's Heisenberg uncertainty principle would be an example of this – it states that it is theoretically impossible to simultaneously determine the current precise state of both the position *and* momentum of elementary subatomic particles, and so the particle's current mechanical state must have a probabilistic rather than deterministic description). In this setting, probabilistic causation is unavoidable.

A Hierarchy of Causal Relationships?

Given the hundreds of years of philosophical controversy over whether causation exists in any knowable sense, it is perhaps reckless to attempt to rank types of causation by their causal strength. But the above discussion suggests a form of hierarchy of causation may be intuitively possible, which is summarised below in order of decreasing strength:

- Necessary and sufficient conditions, C, and laws deduced from self-evidently true postulates (e.g. Pythagoras' Theorem and Euclidean geometry) together deductively imply a certain effect, E. This we might call deductive knowledge.
- Necessary and sufficient conditions, C, and scientific theories that can be regarded as empirically confirmed and hence probably true, at least within some domain of application,

(e.g. Newtonian laws of motion) together deductively imply a certain effect, E. This is deterministic causation.

- Conditions, C, and scientific theories that can be regarded as empirically confirmed and hence probably true, at least within some domain of application, together deductively imply that an effect, E, is probable but not certain, and more probable than in the case of not-C (e.g. theories of genetics and hereditability). This is probabilistic causation.
- Conditions, C, and an empirical generalisation of a deterministic or probabilistic relationship that is not explained (e.g. Boyle's Law prior to the kinetic theory of gases) inductively implies an effect, E. This is an inductive generalisation and is not a form of causal explanation.

2.4 Realism and instrumentalism

Our brief survey of some major topics in philosophy of science next considers scientific realism and its counterpart anti-realism or instrumentalism[148]. Philosophical realism refers to the metaphysical idea that the real world exists independently of the observer. Its philosophical counterpoint is idealism, which contends that nothing exists (or, perhaps, can be presumed to exist) except for the mind of the observer, such that the external world is essentially a mental construct. There are various schools of idealism with differing degrees of commitment to the idealist concept – solipsism is an extreme version that believes that the only thing we can know exists is our own mind; phenomenalism is a variant of idealism that says nothing can be presumed to exist other than what can be reduced to a description of our own sensations. The reader may be relieved to hear that we shall not be attempting a detailed treatment of these metaphysical perspectives. Instead we will focus on realism and its counterparts in the specific context of the scientific method.

Scientific theories will generally postulate a model of reality that is idealised, incomplete and of applicability only within some limited domain. The modeller's clichéd phrase that 'all models are wrong, some are useful' also applies to virtually all scientific theories. Philosophers of science hold different views about the extent to which the *realism* of a theory's premises matter. A scientific realist will generally hold that scientific theories should be interpreted literally (including in their use of theoretical or unobservable entities) as attempted descriptions of reality. A realist would hold that a theory is either true or false, although it may only be possible to determine its truthfulness to a degree of probability (which may or may not be quantifiable), and they may accept the theory is only an attempted *approximation* of reality. So, to the realist, a scientific theory attempts to use true premises to deduce empirically accurate conclusions.

An instrumentalist, on the other hand, would adopt a converse view by arguing that the purpose of a scientific theory is not to provide a description of reality - rather, it is to make useful predictions about a particular range of phenomena that can be empirically tested ('the object served by the discovery of such laws is the economy of thought' in Karl Pearson's words[149]). The instrumentalist outlook can naturally be associated with empirical sceptics. And it has been particularly associated with the *phenomenalist* form of idealism, which, as noted above, asserts that the only form of sure knowledge is that derived directly from our sensory experience and introspection (though a phenomenalist perspective is not necessary to hold an instrumentalist view of scientific theories). This has broadly been the position of the historical titans of British empirical philosophy, such as David Hume and Bertrand Russell (at least over a part of his career)[150], as well as other important

[148] See Ladyman (2002), Chapter 5 for a highly-accessible introduction to this topic; and Nagel (1979), Chapter 6 for a deeper analysis.
[149] Pearson (1911), p. 78.
[150] Russell (1912-1913)

philosophers and scientists who have been concerned with probability and its application to the scientific method, such as Ernst Mach and Karl Pearson[151].

To an instrumentalist, the truthfulness of a theory's assumptions as a description of reality is incidental rather than of fundamental relevance. And so, to an instrumentalist, a scientific theory need not attempt to provide true premises from which to make deductions. The truthfulness of the premises are essentially irrelevant to the instrumentalist. Rather, the purpose of the premises and theory that builds on them is to provides an inferential device or principle which, when combined with some initial conditions, can provide good descriptions of the behaviour of other empirically observable phenomena. In this sense, theories are intellectual tools (or instruments) rather than models of reality. Nagel described the instrumentalist perspective thus:

"A theory is held to be a rule or a principle for analysing and symbolically representing certain materials of gross experience, and at the same time an instrument is a technique for inferring observation statements from other such statements."[152]

Both scientific realism and instrumentalism would seem broadly compatible with the scientific method of both the falsificationist and Bayesian varieties (though an instrumentalist would not attach a probability to the 'truth' of a theory, as the theory's 'truthfulness' is not relevant to the instrumentalist). In the hierarchical structure of scientific theories that was set out by Braithwaite (see Chapter 2.1), the difference between realism and instrumentalism ultimately boils down to whether the scientist regards the theoretical, non-observable terms of the hierarchy's higher-level hypotheses as representative of reality or not. But whichever view is taken has no significant consequence for the scientific method as represented by the structure. Indeed, it could be argued that the differences between realism and instrumentalism are essentially linguistic and involve philosophical hair-splitting.

There is, however, at least one aspect of scientific output that is fundamentally impacted by the question of realism versus instrumentalism: causation and scientific explanation. The definitions of scientific explanation developed by the logical empiricists such as Hempe and Oppenheimer assumed that the scientific laws at the core of scientific explanation were in fact true. Instrumentalists like Mach and Pearson held the view that scientific explanation and the identification of causation, at least as defined in Chapter 2.3, is simply not something that empirical science can deliver. For realist philosophers like Popper, for whom explanation is one of the fundamental goals of science, this was not an ambition that they were willing to forego.

There are certainly some aspects of scientific practice that support the instrumentalist perspective. For example, a scientist may employ two scientific theories which are incompatible (in the sense that they both cannot be true descriptions of reality) in two different fields of study, because the scientist understands which one provides the most efficient or accurate description of a particular type of phenomena. This gives rise to what Nancy Cartwright, the contemporary American philosopher of science deemed, 'a patchwork of laws'. But, equally, there is significant aspects of scientific practice that appear to be more aligned with a realist perspective. For example, much experimental research is involved in determining whether or not the theoretical entities of a given theory actually exist. A purely instrumentalist perspective arguably would not care.

A realist could argue that the success of science in predicting and explaining phenomena is strong testament to the ability of scientific theories to get to the truth of how the world really works. This is

[151] Pearson (1911)
[152] Nagel (1979), p. 129.

sometimes referred to as the 'no-miracles' argument – that is, it would be rather miraculous for scientific theories to obtain any degree of success if they did not actually approximate reality reasonably well. An instrumentalist might counter that the radical (and incompatible) way that empirically successful scientific theories of a given phenomenon have changed over historical periods of time suggests it would be complacent to assume today's scientific theories are reliable descriptions of reality (philosophers of science have referred to this argument as 'the pessimistic meta-induction'[153]). But this is not necessarily inconsistent with the realist outlook that the evolution of scientific theories represents an ever-improving approximation to reality and an objective idea of truth (this would be broadly representative of Popper's realist perspective).

100 years ago, the instrumentalist perspective held sway. Even as late as the 1930s, Popper's scientific realism was seen as a major departure from the prevailing orthodoxy. However, in contemporary philosophy of science, some form of scientific realism is now the orthodox position, with a few notable dissenters (such as Bas Van Fraassen[154]). This historical shift from instrumentalism to realism is perhaps ironic given that the most significant breakthrough in scientific theory during this time (quantum theory) seems incapable of realist physical interpretation by either scientists or philosophers. Despite the prevailing philosophical orthodoxy of scientific realism and the unquestioned scientific success of quantum theory, the ambiguity in the physical interpretation of its equation means it is nonetheless the case that "many physicists find it most sensible to regard quantum mechanics as merely a set of rules that prescribe the outcome of experiments"[155].

In a nutshell, the instrumentalist perspective is primarily interested in the performance of the theory in making useful predictions; but it has hard to see how a scientist can develop a model that can be expected to perform well in making accurate empirical predictions if the model is not intended to capture some important aspects of reality. As noted above, this topic of realism versus instrumentalism may seem to verge on philosophical hair-splitting. But the issues with which it is concerned – whether gaps between theory and reality matter; what the main purpose of a scientific theory is - are fundamental. These issues can be particularly important in the context of the social sciences. The realism / instrumentalism debate may be especially interesting for actuarial science when considering the objectives, usefulness and limitations of economic and financial models and theories. The topic will therefore recur in Chapters 3 and 4.

2.5 Post-positivism

Much of this chapter's discussion has been derived from the quarter-century or so of extraordinary progress in the philosophy of science that occurred between between the mid-1920s and mid-1950s. During this period, which featured the logical positivists and logical empiricists as well as other highly influential thinkers such as Popper, the recognised modern model of the positive scientific method was established. Popper's falsificationism (and realism); Braithwaite's model of the hierarchical hypothetico-deductive structure of scientific theory, accommodating both theoretical and observable terms; Hempel's deductive-nomological model of scientific explanation; all had been developed by the mid-1950s as pillars of a philosophy of science that established science as objective and continuously progressive.

We saw above how falsificationism and Bayesianism both implied a scientific method focused on rigorous empirical testing, that values simplicity, and that holds scientific knowledge to be fundamentally tentative and fallible rather than certain. Both philosophical outlooks assumed

153 Laudan (1981)
154 Van Fraassen (1980)
155 Salmon (1984), p. 248.

science is conducted by scientists behaving rationally and objectively in the sense that two scientists would be expected to conduct themselves in the same way and to reach the same conclusions, irrespective of their personal predilections or social context. The implication of this perspective for continuous scientific progress is inherently positive. New theories will be improvements on old theories - they will be more accurate, more testable, or simpler, or have more generality, whilst still containing the information of the old theory. To a realist, science will provide an ever-improving, closer approximation to objective reality and truth. To an instrumentalist, science will provide reliable empirical predictions of ever-improving accuracy and scope.

It was noted, however, that even in this established logical, objective framework for scientific progress, limitations to testing of 'compound' hypothetical structural systems could lead to some potential for subjectivity or need for judgement in drawing conclusions from empirical testing. Chapter 2.1 noted how the Duhem-Quine thesis introduced a complexity to this process of continuous testing and refinement, particularly for advanced sciences. There, the multiple hypotheses that would be contained in a given scientific theory created a need for some form of professional judgment or even hunch-making on the part of the scientist when deciding which hypothesis to reject in the face of some falsifying evidence. Moreover, in this scenario, there may be a natural and arguably a logical inclination to choose to reject the hypothesis that was the least disruptive at that point in time for the overall deductive system of the particular field of science. We discussed how, when taken to its logical limit, this could mean that some hypotheses may become essentially unfalsifiable or 'functionally *a priori*' (at least, until further evidence is gathered).

In the 1960s and beyond, some philosophers of science started to challenge this (predominantly) sunny perspective on the positivist scientific method. This did not lead to the emergence of a new consensus in the philosophy of science and scientific method. Rather, this 'post-positivist' era is better recognised as one where a series of philosophers and schools identified some potential limitations or caveats that ought to be recognised in the positive model of science that was established by the end of the 1950s.

Popper, the logical positivists and logical empiricists were all primarily concerned with developing a logical philosophical framework that could provide a model of how science *should* be done. This work made little direct reference to the actual practice of scientists. The 'post-positivist' era of contemporary philosophy of science has paid much more attention to describing and explaining how and why scientists 'do' science. Post-positivist philosophy of science therefore inevitably tends to touch more on social sciences such as sociology and psychology; and it may seek its confirmation directly from episodes in the history of science.

Kuhn: Paradigms, Revolutions, Incommensurability and the Growth of Scientific Knowledge

One of the major themes of post-positivist thought has centred on the 'growth of scientific knowledge' and, in particular, whether the orthodox positivist view of science as an enterprise that makes relentless forward progress was too simplistic and optimistic. Thomas Kuhn, an American historian and philosopher of science, made a major contribution to the 'growth of knowledge' debate with his ground-breaking 1962 book *The Structure of Scientific Revolutions*[156]*.* This work might be considered as the first major break from the logical empiricist orthodoxy described above. Kuhn considered more fully the complexities that arise when scientists obtain falsifying evidence for advanced scientific systems. He viewed Popper's characterisation of falsification as a mere 'methodological stereotype'. Kuhn proposed an alternative model of scientific progress that paid particular attention to the (current and historical) working practices of scientists and the social and

[156] Kuhn (1996)

cultural norms and incentives that can be associated with them (one might say that, in a way very distinct from Popper or Jeffreys, Kuhn was writing about the *sociology* of scientific communities[157]).

Kuhn argued that scientific activity in the real world could be sorted into two distinct categories: normal science and revolutionary science. Normal science, which Kuhn argued was what the vast majority of scientists spent most of their time on, was concerned with incremental improvement and refinement of well-established scientific theories. Revolutionary science, on the other hand, was concerned with displacing and replacing major theoretical systems. At the core of this distinction was the idea of a 'paradigm'. Kuhn's wide-ranging and inconsistent use of the term attracted some criticism, but a paradigm was essentially "the entire constellation of beliefs, values, techniques and so on shared by the members of a scientific community". Normal science was concerned with working within the paradigm, refining it and developing its applications. Revolutionary science was work that challenged the accepted beliefs of the paradigm and ultimately could lead to the replacement of the existing paradigm with a new one. Importantly, Kuhn argued that the new paradigm may be fundamentally inconsistent with the old one (in ways to be further discussed below).

In the context of the above discussion, one of the most important points of Kuhn's argument is that the social and cultural norms and incentives that scientists work amongst may reduce the objectivity of scientific output. Most notably, Kuhn argued that scientists may be reluctant to give up on paradigms easily, even when have been objectively falsified. This resistance may partly be a practical matter – there is arguably little point in rejecting a theory until a new and better one has been adequately developed. More interestingly, Kuhn argued that the leaders of the scientific community would be incentivised to resist the rejection of the paradigm in which they are experts. After all, rejection of the paradigm would imply that their work and expertise were becoming irrelevant and would be superseded by that of others. Thus, paradigm-shifts may take time (a generation or longer) and may involve significant resistance and professional controversy as they occur[158].

The notion that scientists may become irrationally or dogmatically attached to the particular groups of theories that they were experts in was probably not an especially new or controversial one to working scientists. For example, Max Planck, the theoretical physicist most associated with the original development of quantum theory, and hence intimately familiar with revolutionary science, wrote prior to his death in 1947:

"An important scientific innovation rarely makes its way by gradually winning over and converting its opponents: it rarely happens that Saul becomes Paul. What does happen is that its opponents gradually die out and that the growing generation is familiarised with the idea from the beginning."[159]

As is so often found in the history and philosophy of science, big ideas have a long and complicated provenance. Kuhn, however, placed Planck's common-sense view of practising scientists within a formal philosophical framework and explored its logical consequences.

---

157 We might go even further and say that Kuhn's focus on historical case studies was aligned to the interpretative research methods of Max Weber (see Chapter 4.3).

158 The introduction of risk-based capital techniques to the British actuarial profession in the 1970s and the decade-long controversy spanning the period from the presentation of Benjamin's first paper in 1971 to the publication of the Maturity Guarantees Working Group report in 1980 could perhaps be considered an example of a paradigm-shift within the actuarial profession. See Turnbull (2017) Chapter 5 for further historical details of the episode.

159 Planck (1959), pp. 298-99.

Kuhn further argued that when a paradigm-shift eventually does take place (which could be decades after the first falsifications of the old theory are documented), the theories of the new paradigm could be so different to the previous one as to be 'incommensurate' with it. This means the terms of the theories of different paradigms may not be mutually translatable – for example, 'mass' in Newtonian mechanics means something different to 'mass' in Einstein's theory of relativity (though it may be argued that incommensurate higher-level hypotheses need not lead to incommensurate lower-level hypotheses). Incommensurate theories may not be directly comparable. It may be the case that the new paradigm provides new or better explanations for some empirical phenomena, but at the same time is not capable of explaining some of the empirical behaviours that the previous paradigm explained well. Consequently, the old theory cannot always be merely regarded as an inferior special case or approximation of the new one.

This notion of incommensurability is at odds with Popper's vision of new theories as incremental improvements on old theories where the new theory offers something better and nothing worse. Popper readily conceded that the new theory may provide only a 'good approximation' to some of the valid empirical content of the old theory. But this contrasted with the view expressed by some post-positivists (such as Feyerabend, who is discussed further below) who argued that history demonstrated that the overlap between the old and new theories actually tended to be quite small. This implies that scientific knowledge may not always be cumulative and that scientific progress may not inevitably move in a uniformly forward direction. Instead, a new theory, although an improvement in at least some respects over the old theory, may ultimately prove to have been a wrong turn. Moreover, it could take generations to reverse back out of a paradigmatic cul-de-sac.

Lakatos: A Third Way?

The work of Imre Lakatos can be viewed as an attempt at a middle ground between the philosophies of Kuhn and Popper. Lakatos proposed an alternative model to Kuhn's paradigms and the bipolar world of normal and revolutionary science in his characterisation of a 'scientific research program'[160]. Lakatos' research programs attempted to recognise some of the conventions of scientific communities (in the form of negative and positive heuristics which determined the type of activities that were acceptable within a community), whilst at the same time rejecting the potential for subjectivity and social psychology that characterised Kuhn's perspective.

Lakatos' scientific research program was a group of interrelated theories. The research program had a 'hard core' of key assumptions that were regarded by the scientific community as essentially irrefutable. When falsifiying empirical results are produced, it is the auxiliary assumptions outside the hard core that are changed first. The program would reach a crisis when the changes required to the auxiliary assumptions became 'degenerative'- that is, the changes to the assumption were essentially *ad hoc*, and failed to provide any new empirical content.

Lakatos' scientific research program can be viewed as broadly consistent with Kuhn's description of scientific practice: Kuhn's paradigm and Lakatos' hard core share obvious similarities, as do Kuhn's revolutions and Lakatos' crises. But the scientific research program can also be viewed as being consistent with Braithwaite's hierarchical structure of scientific theory as a model of the logical relationships between a large system of interrelated theories. Decades before Lakatos, Braithwaite had argued that it was logical that a system of theories would be altered in the least disruptive way when faced with falsifying results (see Chapter 2.1 and the section on *Falsification of Systems of*

[160] Lakatos (1970)

Theories). And Pap's 'functionally a priori' higher-level hypotheses – the highest levels of Braithwaite's hierarchy - sound very similar to the scientific research program's hard core.

Braithwaite and Pap's arguments did not require any appeals to sociology or psychology. Nonetheless, both the Braithwaite-Pap and Lakatos perspectives provide an explanation of how science was conducted as a rational methodological system that promoted the positive growth of scientific knowledge in a way that was more complex than the 'naïve falsificationism' originally advocated by Popper. Perhaps both perspectives can offer some insight into the historical and current working practices of scientific communities. After all, they are not mutually exclusive. Reluctance on the part of a scientific community to reject its scientific paradigms or the hard core of its scientific research programs may be both an inevitable logical consequence of the Duhem-Quine thesis in the sense discussed by philosophers such as Braithwaite and Pap; and, at least in part, also a natural sociological consequence of the vested interests that the leaders of a scientific community inevitably have.

Feyerabend's Radical Contribution to Post-Positivism

Some other philosophers of science of the post-positivist era have been notably more radical than Lakatos in their rejection of the key tenets of the positive empirical scientific method. Perhaps the most radical of these philosophers is Paul Feyerabend. Feyerabend collaborated with Kuhn at Berkeley in the late 1950s. There is some significant consistency in their views, but Feyerabend is undoubtedly more revolutionary in his perspectives and prescriptions. His most influential book, *Against Method*, was first published in 1975. There, Feyerabend argued that, in the real world, scientific activity was too idiosyncratic and complex to be well-represented by any philosophical model of scientific method. He therefore argued for a sort of methodological anarchy: *anything goes*.

We noted earlier that some have challenged the idea that observations can be completely detached from the theory which used them. Feyerabend advocated an extreme version of this 'theory-dependence thesis'. He argued that the meanings of virtually all forms of scientific observation are highly dependent on the scientific theory that uses those observations: "the interpretation of an observation language is determined by the theories which we use to explain what we observe, and it changes as soon as these theories change."[161]

A corollary of this view of observation as deeply theory-laden is that alternative (and incompatible) theories may be necessary merely to illustrate that a given theory is refuted by existing observational data[162]. Thus, to Feyerabend, the regimented logic of falsificationism, where one theory reigns supreme until it is shown to be false, is an anathema – according to both Feyerabend's philosophy of knowledge and his interpretation of the history of science, science prospers in a state of chaotic competition amongst incommensurate alternative theories, all of which are capable of being falsified by some interpretation of the available evidence ("A strict principle of falsification…would wipe out science as we know it and would never have permitted it to start."[163]).

Theory-laden observation makes the incommensurability of alternative theories inevitable (theories' predictive performance relative to some empirical observations cannot be usefully compared if the observations themselves are given different interpretations by the respective theories). But Feyerabend regarded incommensurability as an over-rated philosophical construct with little real

[161] Feyerabend (2010), p. 219.
[162] See Feyerabend (2010), p. 12 and p. 22.
[163] Feyerabend (2010), p. 157 (see also p. 45).

scientific consequence: “Incommensurability disappears when we use concepts as scientists use them, in an open, ambiguous and often counter-intuitive manner. Incommensurability is a problem for philosophers, not for scientists.”[164]

Whilst Feyerabend’s ideas and arguments were never widely accepted by philosophers or scientists, they nonetheless have played a useful and influential role in challenging the potential over-simplification, dogmatism and complacency that could arise from the substantial development of (a positivist and deeply empirical) philosophy of science that took place earlier in the twentieth century.

[164] Feyerabend (2010), p. 219.

3. On the methodology of the social sciences

Chapter 2 highlighted some of the diverse philosophical topics and perspectives that are relevant to the methodology of science. These discussions were set in the context of the natural sciences. In the domain of these sciences, the subjects of study are natural, physical phenomena and their observation can usually (though not always) be undertaken in carefully controlled and repeatable experimental conditions. A principle of the uniformity of nature through time and space permits scientific knowledge to be built through some form of inductive inference from the observed to the unobserved. Under these conditions, quantitative probabilistic or statistical inference, where required, is often feasible due to the presence of data samples that are explicitly well-behaved (independent and identically distributed in the jargon of probabilistic objectivists; exchangeable in the jargon of probabilistic subjectivists).

This chapter re-visits philosophical questions of methodology when working in the domain of the social sciences. The essential distinction between the social and natural sciences is that the social sciences are concerned with the study of some form of *human* behaviour rather than the behaviour of physical phenomena. The social sciences may encompass any field of study that allows theories of some aspect of human behaviour to be developed and tested empirically. Subjects that could feasibly be defined as a social science include economics, psychology, anthropology, sociology and politics. A further distinction can be made between subjects that are about the behaviour of the individual human (for example, most of psychology) and subjects that are concerned with the collective or aggregate behaviour of humans (for example, economics or sociology).

The introduction of a human behavioural element to the subject of scientific study brings a raft of potential complications to its methodology. These can be roughly grouped under three headings: the free will of human beings; the non-stationarity of human societies; the difficulties often associated with controlled observation in the social sciences. Let's briefly outline each of these topics:

- *Free will*. Can we theorise, generalise, explain and make predictions about the (individual or aggregate) behaviour of conscious, self-determining individuals that possess free will? Should we aim to scientifically study how humans actually behave, or how they would behave if they had a particular set of motivations or predispositions? If the latter, how can the theory be empirically tested? Is there a set of such motivations that can be taken as self-evident axioms of social science? Could theories of human behaviour become self-fulfilling or self-defeating in the sense that the theory itself may alter the human beliefs, expectations and actions that the theory is intended to explain or predict.
- *Non-stationarity*. The characteristics of human societies change over time in ways that have no obvious parallel with, say, the characteristics of natural gases. The applicability of a principle of uniformity of nature over time and space may therefore be harder to justify. This makes prediction profoundly more difficult. In the natural sciences, the time at which a repeatable experiment is conducted is, by the definition of repeatability, generally irrelevant (clearly, there are important exceptions where repeatable conditions may not be so easily observed, such as in the fields of astronomy or geology). But in the social sciences, the non-stationary nature of some of society's conditions may mean that the time of an empirical observation of some aspect of human behaviour may be of fundamental importance.
- *Observation*. Even if human society could be described as stationary, conducting observation by controllable, repeatable experiment may have many inevitable practical difficulties in the social sciences. For example, it would take a very long time to reliably observe how different monetary policies impact on the long-term rate of inflation.

These methodological hurdles have arguably resulted in the social sciences emerging and progressing more slowly than their natural science cousins. We might date the emergence of modern natural science to the Copernican revolution of 1543. A comparable event in the early history of the social sciences could be taken to be the publication of Adam Smith's *Wealth of Nations* over two hundred years later in 1776. Whilst the methodology of natural science has been philosophically debated since Francis Bacon's *Novum Organum* of 1620, discussion of methodology in social science had to wait until the mid-nineteenth century and the arrival of the works of Mill, Comte and the German historicists.

Whether an advocate of Popperian falsificationism or Kuhnian revolutions; whether a frequentist or a Bayesian; a realist or an instrumentalist; virtually all philosophers of science would agree that a (or perhaps *the*) defining feature of scientific method is its empiricism; science is about explaining natural phenomena by general theories, models or laws, and providing observational evidence to support how well these theories usefully represent regularities that can be empirically observed. This fundamental feature of science and its method should be true of the social sciences as well as the natural sciences; indeed, it is what *makes* a social science a science rather than, say, one of the humanities. Given the complicating factors set out above, can this be done? Can there be a unity of scientific method that is applicable across the natural and social sciences, and for which there can be a reasonable expectation of success in the social sciences? Many philosophers of science have argued that it can be. Or, at least, to the extent that empirical knowledge can be developed in the social sciences, it should be acquired through the application of the scientific method that has been successfully applied to natural science. As Karl Pearson, writing at the turn of the twentieth century, put it:

"The scientific method is the sole gateway to the whole region of knowledge. The word science is here used in no narrow sense, but applies to all reasoning about facts which proceeds, from their accurate classification, to the appreciation of their relationship and sequence."[165]

As we will find below, this perspective is not one that is unanimously held amongst philosophers. But let us begin under the premise that it is at least possible to acquire scientific knowledge in the fields of the social sciences. Questions then naturally follow: how do the special features of the social sciences impact on its methodology? Does it require methodological approaches that are completely distinct and with no parallel in the natural sciences? And is the nature of its output likely to be different or more limited than the output of the natural sciences? For example, will the scientific generalisations of the social sciences tend to merely describe some recurring empirical relationships, or can they take the form of theories that have deductive and testable implications that can form larger logical systems of scientific explanation? Do the social sciences generate scientific knowledge of a similar form to the scientific knowledge of the natural sciences? This chapter considers questions such as these.

3.1 History as a form of scientific explanation

The subject of history is not usually regarded (by historians or philosophers) as meeting the criteria required to be classified as a social science. History may therefore appear an odd place to start in our discussion of methodology in the social sciences. However, drawing out the distinction between methodology in history, and methodology in, say, sociology, which also uses historical events as empirical data for its studies, can help to develop a fuller understanding of the demarcation between the humanities and the social sciences.

[165] Pearson (1911), p.24.

The fundamental objective of the academic study of history is to understand, reconstruct and explain the unique and specific features of a *singular* historical event or series of events in as much detail as possible. This sharply contrasts with the basic scientific activity of making universal generalisations from a set of many controlled, repeatable observations. As we noted in Chapter 2, scientists view simplicity as a fundamental desideratum in their descriptions of natural phenomena. Historians, on the other hand, value complexity in theirs: 'historians are as remorseless in their pursuit of ever more microscopic explanations as sociologists are zealous in their pursuit of ever broader laws'[166].

The prevailing view of professional historians is that the primary purpose of the study of history is to understand the past and the path that has been trodden to the present time because it is of some intrinsic interest to us – it is not to develop inductive inferences form 'historical lessons' that teach us that a particular set of circumstances will always (or usually) be followed by another. As Lionel Robbins, the great philosophically-minded economist of the early 20th century noted in his critique of historical induction, "It is one of the great merits of the modern philosophy of history that it has repudiated all claims of this sort, and indeed makes it the *fundamentum divisionis* between history and natural science that history does not proceed by way of generalising abstraction."[167]

Nonetheless, some historians and philosophers of history have argued that it *is* possible to develop causal explanations for what has happened in the past. That is, to identify the historical effects that are followed by the presence of specific causes. Can such a causal historical explanation be *scientific*? How can such an explanation be empirically tested? Can the notion of a 'microscopic explanation' of a unique historical event fit naturally with the definitions of scientific explanation that were discussed in Chapter 2.3? There, explanation made use of universal (or statistical) generalisations about the behaviours of some form of phenomena – what are sometimes called laws of nature. Can similar forms of causal laws be identified that can deliver scientific explanations for specific events of history? Interestingly, Carl Hempel, the German-American philosopher of science and logical empiricist who was so influential in the development of philosophical thought on scientific explanation, argued this was indeed the case.

The Covering-Law Model of Historical Explanation

The long-standing philosophical debate on the plausibility of historical explanation was re-animated in the middle of the twentieth century when Hempel turned his attention to the methodology of history. In his 1942 essay, *The Function of General Laws in History*[168], he argued that causal laws *could* be used to explain the occurrence of historical events using similar logical structures of explanation as he advocated for explanation in the natural sciences. Moreover, he argued that these laws (which may be deterministic but would more likely be probabilistic in form) would be as scientifically legitimate as those produced in the natural sciences. Laws used in historical explanation may, according to Hempel, be directly derived from historical tendencies that were identified through the categorisation and analysis of disparate historical events (for example, in this way we might identify the conditions that tend to make political revolutions more likely). Hempel's proposal that scientific laws could be identified which could provide explanations for historical events became known as *the covering-law model of historical explanation*.

Hempel's argument ran contrary to history's conventional humanist outlook and triggered much philosophical debate through the 1950s and 1960s. The explanatory power of historical analysis was

---

[166] Hempel (1959)
[167] Robbins (1935), p. 77.
[168] Hempel (1942)

strongly rejected by Karl Popper in his 1957 book *The Poverty of Historicism*[169]. Popper argued that every historical event is unique and infinitely complex, and that societies are continually changing, making such comparative historicist analysis futile - the categorisation of like events that Hempel envisaged is not possible if no two events can be known to be alike. It makes the discernment of historical cause and effect impossible:

"Long-term prophecies can be derived from scientific conditional predictions only if they apply to systems which can be described as well isolated, stationary and recurrent. These systems are very rare in nature; and modern society is surely not one of them."[170]

Max Weber, the influential early 20th century social scientist, had similarly argued that reality is "an infinite causal web" and that "the reality to which the laws apply always remain equally individual, equally undeducible from laws"[171]. And before him, Venn also argued in the late 19th century that the world was infinitely complex and that any historical event and its circumstances must therefore be unique. But Venn nonetheless argued that recurring features may be identified in historical events "which may offer close comparison with what we may again experience."[172]

This important concession of Venn's plays an important role in Hempel's theory of historical explanation. Hempel's covering-law argument did not go so far as to argue that his explanatory laws must be deterministic – as the above Popper reference to 'prophecies' alludes, this would amount to being able to perfectly predict the future as well as perfectly explain the past, which seems rather fanciful, even for the most ambitious project of logical positivism. Rather, Hempel argued that these laws would tend to be *probabilistic*. He argued that many historical explanations would take a similar form to an example such as: "if Tommy comes down with the measles two weeks after has his brother, and if he has not been in the company of other persons having the measles, we accept the explanation that he caught the disease from his brother."[173]

From this perspective, Hempel's arguments seem less ambitious and controversial. It is difficult to argue with the idea that we can form rational judgements on the sort of general historical circumstances that may make some form of specific outcome subsequently more likely. Medicine, law and most other things in life must rely on the use of this form of judgement. Whether we consider such singular, probabilistic judgements as forms of scientific knowledge of a similar form to the scientific knowledge of the natural science is perhaps a matter of epistemological taste.

Determining the Empirical Content of Laws of Historical Explanation

The fundamental test of whether historical explanation can be scientific hinges on the empirical testability of the causal laws that are proposed in the explanations. As we saw in Chapter 2.3, Hempel's own definition of an acceptable scientific explanation required the use of a law that had *empirical content*. How can the observation of singular historical events, each with their own unique set of circumstances, be used to form testable explanations of cause and effect from which such content can be derived?

Ultimately, inductive evidence, or empirical content, in support of a historical causal law must be based on the identification of patterns in historical events. One of the simplest ways of identifying

[169] Popper (1957)
[170] Popper (1959)
[171] Weber (1949), p. 73.
[172] Venn (1889), p. 395.
[173] Hempel (1959), p. 350.

such patterns is by what John Stuart Mill called *the Method of Difference*. This method was introduced as one of his canons of induction in *System of Logic*:

"If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon."[174]

Note this is really a piece of deductive logic. So why did Mill refer to it as a canon of induction? Because it may not be possible to *know* if all the circumstances but one are in fact entirely identical. After all, history cannot be re-run with the (alleged) causal factor removed. So, in the context of explaining historical events in all the complexity of the real world, and without the possibility of highly controlled experimental conditions or access to counter-factual observations, such reasoning cannot be exact and deductive, and so is reduced to a form of *inductive* argument.

Consider, to take one of many possible historical examples, Weber's 1905 book *The Protestant Ethic and the Spirit of Capitalism*[175]. Its thesis was that the emergence of Protestantism was the crucial factor in the development of capitalism in 18th century northern Europe. The logical argument for this claim can be made by arguing that other historical and geographical examples can be identified that had the same necessary pre-conditions for the growth of capitalism as 18th century Europe other than the presence of Protestantism, and yet did not experience the same capitalist development. Expressing this argument in the formal structure of Chapter 2.3, we can identify some number, say, *r* observations which share the same conditions C_1, C_2,…C_n and all these observations are associated with the absence of the effect of the growth of capitalism. We then have another observation, which is the period of northern European history when Protestantism emerged, which shares the same circumstances C_1, C_2,…C_n as the *r* observations, but where C_{n+1}, the emergence of Protestantism, is also present.

This form of argument can be seen as an example of Mill's *Method of Difference.* It leads to the conclusion that the emergence of Protestantism was a necessary condition for the development of capitalism. But are the premises of this argument reliable? How do we know that these (n+1) conditions form an exhaustive list of the circumstances necessary and sufficient for the development of capitalism? How can we be sure that the conditions C_1, C_2,…C_n are all present in exactly (or sufficiently?) similar form in all (r+1) observations? Incidentally, historians have disputed the observational basis of Weber's argument by noting that many late Medieval European cities of predominantly catholic faith such as Venice and Florence appeared to have engaged in a highly successful form of capitalist society.

For completeness, it should also be noted that Weber did not explicitly use this form of logical argument in the book – he focused largely on explanations for *why* the cause it existed rather than *if* it existed. The determination of the empirical content of the assumed historical causal law was not a feature of the book. Instead, he was more concerned with the idea that the Calvinistic dogma of predestination played a particular role in transforming the meaning of a man's labour from a mundane necessity into a form of religious calling, spiritually compelling him to work hard for the glory of God as 'the only means of attaining certainty of grace'[176].

[174] Mill (1844), p. 455.
[175] Weber (1930)
[176] Weber (1930), p. 106.

The uniqueness of historical circumstances, and the lack of comparability that follows from it, presents a fundamental difficulty in the development of a scientific form of historical explanation. The discussion of scientific explanation in Chapter 2.3 emphasised that, for a condition to be deemed explanatory of some effect, the effect must be absent when the condition is absent, if all other conditions remain unchanged. If we cannot use different historical scenarios to make *ceteris paribus* comparisons of the effect of the presence or absence of a cause, there would seem to be no other way of isolating cause and effect. As noted above, in a given historical scenario, we cannot directly observe not-C – it is a counterfactual scenario, by definition it does not exist and cannot be observed. In these circumstances, a historian or social scientist may explain why C is the cause of E by providing an argument for what can be inferred would occur (i.e. not-E) in the event of not-C. But such an inference will tend to rely on an assertion about, for example, a person's intentions or disposition or some empirical generalisation about human behaviour[177]. *It is impossible to test directly whether the assertion is correct within the unique circumstances in which it is being applied.* As the historian A.J.P. Taylor once quipped: 'How can we decide about something that did not happen? Heaven knows, we have difficulty enough in deciding what did happen.'[178]

Mill's a priori *solution to the lack of empirical content in historical causal laws*

In his late-19th century discussions of historical explanation, J.S. Mill wrote of the 'certain general tendencies' (as opposed to the rigorous regularities) that could be identified in an empirical analysis of historical events. He argued that this lack of precision in observable historical relationships was because 'properties are changeable' (that is, to a statistician, the historical process is non-stationary) and he further argued that this non-stationarity arose primarily because, in society, effects feedback into causes in complex ways[179].

Mill did not, however, conclude that this made the empirical analysis of historical relationships worthless. He argued that the relationships identified in empirical historical analysis could not be regarded as causal laws of nature, but rather as derivative laws that were driven by more fundamental (stationary) general laws of human behaviour that were then combined with the (non-stationary) societal conditions of a given point in time. These fundamental laws were not directly observed from empirical observation, but were instead theoretically deduced *a priori* from first principles, i.e. they were taken to be self-evident and axiomatic rather than established by empirical observation. The derivative empirical laws implied by these fundamental theories could then be tested against the relevant historical data.

Mill even (implicitly) deployed a form of Bayesian logic to explain this process: he supposed there should be a high prior probability attached to the derivate empirical laws that have been deduced from the axiomatic fundamental laws of human nature, and that this probability should increase to near certainty when confirmed by positive observation of their empirical consequences:

"Although these results, obtained by comparing different forms and states of society, amount in themselves only to empirical laws; some of them, when once suggested, are found to follow with so much probability from general laws of human nature, that the consilience of the two processes raises the evidence to proof, and the generalisations to the rank of scientific truths…"[180]

This argument is notable for proposing that historical explanation *can* be empirically confirmed, and to a very high degree. But what are these self-evident fundamental unchanging laws of human

---

[177] See Brown (1963), Part Two, Methods of Explanation.
[178] Taylor (1969)
[179] Mill (1959), p. 86.
[180] Mill (1959), p. 90.

nature laws? And how are they combined with specific societal conditions to derivative empirical laws that describe testable human behaviour? It is an interesting model, but it seems to be at best a highly abstract description of how to develop an explanatory theory of human behaviour in a continually changing world. Given the entire point is to deliver empirically testable causal laws, the abstract nature of Mill's structure of explanation means its usefulness is not so obvious.

Genetic Explanation

Interest in a more elaborate form of structure for historical explanation, called a *genetic* explanation, was re-ignited in the 1990s by Clayton Roberts, a leading American academic historian. Hempel had discussed the idea of genetic explanation[181] and Roberts attempted to further develop it, defining it as 'tracing the causal sequence of events leading up to a historical event, with the purpose of explaining the occurrence of that event'.[182]

Roberts argued that, through this analytical tracing process, it would be possible to identify the web of prior events that caused the event of interest. In this model, causal laws would be required to show how each event caused the next in the chain. These laws, however, would not be grand universal descriptions of broad categories of historical events of the kind originally envisaged by Hempel's covering-law framework. Rather, by breaking down the causal chain into its smallest possible steps, the causal laws linking the events could become self-evident common sense:

"Each step in the complicated web of events leading to the war or revolution or reformation has to be connected with an earlier event by a covering law. And if there is a law, must there not be some theory to explain that law? There well may be, but often the covering law is so platitudinous that the historian feels no need to find a theory to explain it."[183]

This perspective on historical explanation is doubtlessly more attuned to the actual activity of historians than Hempel's covering law model. As noted above, historians are almost always interested in describing a historical event in as much detail as possible, rather than attempting to categorise historical events into broad categories that can facilitate sweeping generalisations. Other first-rate scholars have also argued in favour of a form of probability-based historical explanation that is centred on causal sequence. Sir John Hicks' 1969 book, *A Theory of Economic History*, is another example. Hicks argued that study of past historical events would allow the development of a theory of the 'normal development' of an economy through various phases such as customary, command, feudal, mercantile, city state, colonial and industrial.

Roberts argued that his causal laws were so incremental as to be 'platitudinous'. It might be inferred that such self-evident laws were therefore certain, and hence would provide a deterministic and complete explanation of events. But, as noted above, perfect historical explanation also implies perfect foresight, which is generally accepted to be absurd. Roberts avoided this implication by arguing that his self-evident causal laws were merely probabilistic. It is not necessarily easy to envisage what these self-explanatory probabilistic causal laws look like.

The logic of breaking down the causal web of converging events into the simplest, smallest, incremental steps would suggest that the number of links in the chain, or strands of the web, could be enormous. The product of such a volume of probabilistic inferences seems generally unlikely to

---

[181] Hempel was not the first or only writer to embrace the concept of a historical explanation that is constructed by decomposing a causal sequence into a sequence of small steps, each of which that can be empirically tested or viewed as self-evident. For example, see Nagel (1979), p. 568; Weber (1949), p. 173.
[182] Roberts (1996), p. 36.
[183] Roberts (1996), p. 159

result in the generation of strong statements of explanation or prediction. It is distinctly reminiscent of Pascal's ironic tale of how the length of Cleopatra's nose caused the fall of the Roman Empire.

Historical Knowledge, Scientific Knowledge and Epistemic Probability

When Mill, Hempel, Roberts and Hicks refer to probability in the context of historical explanation, they clearly do not mean probability according to the objective definitions of Popper, Von Mises or Venn that were discussed in Chapter 1. Probability here does not refer to relative frequencies found within infinite sets of occurrences. Rather, probability is used here to denote a degree of belief in the epistemic sense, as an expression of incomplete knowledge as used by Keynes, Jeffreys and logical positivists such as Carnap. We find the logical positivists and Popper on their familiar sides of the debate on the validity and usefulness of the fundamental philosophical concept of an epistemic probability of a singular event – to Popper and other probabilistic objectivists, it is an ill-conceived, ill-defined notion; to a positivist such as Hempel, it fits naturally into their philosophy of knowledge.

Irrespective of philosophical perspectives, in the context of the analysis of historical evidence, which by its nature will be highly qualitative, the difficulty of producing 'a generally accepted, explicitly formulated, and fully comprehensive schema for weighing the evidence for any arbitrarily given hypothesis so that the logical worth of alternate conclusions relative to the evidence available for each'[184] is a formidable hurdle to quantifying such epistemic probabilities. This challenge of determining methods for the quantification of epistemic probabilities in the presence of diverse and complex information is familiar from the discussions of Chapter 1. Nonetheless, it could be reasonably argued that we would be throwing the baby out with the bathwater if we refused to concede that some generalisations in the behaviour, motives or tendencies of people can be identified and used to infer likely explanations for historical events. As Comte put it: 'All historical positions and changes must have at least some grounds of justification; otherwise, they would be totally incomprehensible, because they would be inconsistent with the nature of the agents and of the actions performed by them'[185].

Crucially, however, the lack of a basis for the empirical testing of these explanations means that, whilst the explanations may be rational, analytical, objective and logical, they are not scientific, at least not in the empirical sense (is there any other?). The counter-factual scenario – history's crucial experiment – is not generally empirically available. Ultimately, historical explanation may make use of logic, analysis and intellectual rigour, but its empirical limitations puts it beyond the grasp of the scientific method. It therefore should not be considered as a scientific form of explanation. It is valuable and valid, but it delivers a distinctly different form of knowledge to that produced by the scientific method.

3.2 Positivism in the social sciences

There is a major school of thought in the social sciences that attempts to apply the positive empirical scientific method of the natural sciences (as discussed in Chapter 2) to the study of human life, behaviour and society. In social sciences, positivism aims to identify, empirically measure, predict and explain empirical relationships between social phenomena.

The founding of the positivist school of sociology, and perhaps social science writ large, is usually credited to the nineteenth century French philosopher Auguste Comte. Indeed, some date the birth of social science as a 'science' to the publication of Comte's vision of positivism in the 1850s (earlier historical events such as the Reformation and the freedom of thought that it permitted; together

---

[184] Nagel (1959), p. 385.
[185] Comte (1912), p. 20.

with the invention of the printing press and the wider circulation of knowledge that it in turn permitted were clearly important changes in background societal conditions that enabled the study of social science to progress more generally). It was Comte who first coined the term 'positivism' and he was an influence on the Vienna Circle and the logical positivists more widely. The principle of a methodological unity across all forms of science was a key tenet of Vienna Circle members such as Neurath and Carnap. Comte's principal work on positivism is his four-volume *System of Positive Polity*, which was published between 1851 and 1854 (the major English translation was published in 1875-77). His *A General View of Positivism* provides a single-volume 'introduction' to Positive Polity.

To Comte, positivism was not merely a methodology for sociological research. It was the over-arching unifying philosophy of all forms of scientific knowledge and it was capable of delivering not merely knowledge, but immense practical benefit to society, and even spiritual enlightenment. Comte argued that positivism was a stage in intellectual development that only an intellectually sophisticated society could reach. First, society must pass through the necessary but unproductive stages of theology (theological imagination) and then metaphysics (metaphysical discussion) before finally arriving at the permanent intellectual state of positivism (positive demonstration). His objective for positivism was 'to direct the spiritual reorganisation of the civilised world'[186].

Rather like the word 'probability', the term 'positivism' is used widely in philosophy and across the natural and social sciences in various different contexts. In epistemology and the philosophy of the natural sciences, positivism is most often used in the context of the logical positivism of the Vienna Circle that has been mentioned in various places earlier. There, positivism is primarily concerned with developing the application of inductive logic to incomplete information or evidence so as to arise at rigorous measures of probable knowledge, the strength of which may be quantifiable. When used in this sense, positivism may be viewed as more epistemologically ambitious or optimistic than the Popperian philosophy of science, which reflects a narrower epistemology that is closely aligned to the objectivist philosophy of probability (and which rejects the notion of attaching measurable probabilities to uncertain singular events or statements).

In some social sciences such as economics, positivism is most often used as the antonym of normativism – positive referring to how things are; normative describing how things ought to be (and hence including an ethical dimension beyond the factual and empirical basis of science). As will be discussed below, the disentangling of normative views on values and ethics from the positivist activity of establishing empirical facts can be difficult to accomplish in the social sciences.

In sociology and other social sciences today, positivism will usually imply the use of empirical data, and the application of statistical techniques to it, in order to identify, quantify and test empirical relationships within human society in a manner consistent with the generally-recognised principles of the scientific method. This quantitative activity may be used to establish empirical generalisations or to empirically test theories (as well discuss below, the former tends to be more common than the latter in the social sciences). It will often be motivated by a desire to produce knowledge and insight that will support the development of successful public policy solutions (though the choice of policy solution is ultimately a normative, political act beyond the remit of positive social science). From this perspective, some of the earliest work on mortality modelling may be viewed as (a crude precursor to) positivist sociology. For example, John Graunt's 1662 paper was not motivated by the desire to accurately price life assurance policies (indeed, such policies barely existed at the time); rather, it

---

[186] Comte (1912), p. 16.

was explicitly directed at supporting government policy by improving knowledge and understanding of London's mortality rates[187].

So, positivism can be viewed as the philosophical commitment to apply the scientific method that has been so successfully applied in the natural sciences since the scientific revolution to the domain of the social sciences. That is, to quantify empirical relationships between related phenomena; to develop explanatory theories of these relationships; to test the predictions of these theories through further empirical analysis; to continually develop and test new theories that are simpler and / or able to explain more. The introductory passage to this chapter noted some of the difficulties that potentially arise in this undertaking. The relative lack of success in the social sciences in developing empirically tested powerful general explanations of social behaviour that attract the unanimous support of the professional community of social scientists suggests these methodological problems may be significant. These challenges, and their potential solutions, are discussed further below.

Experiment and Observation in the Social Sciences

The ability to observe phenomena behaving under well-defined circumstances is a necessary condition for the practice of empirical science. It is, for a range of reasons, generally harder to conduct such observations in the social sciences than the natural sciences. It will often not be possible to carry out repeatable, controlled experiments in economics or sociology in the way that they usually can be in natural sciences such as physics or chemistry. Causal factors and their effects usually cannot therefore be individually isolated, controlled and measured in the social sciences, but, at best, can only be passively observed. This fundamental difficulty that arises in the practice of empirical social science has long been recognised by philosophers. It was written about by Comte and John Stuart Mill in the middle of the nineteenth century[188] and it has been written about extensively ever since. However, many significant philosophers[189] have noted that a lack of controllability and repeatability in experimental observation is also an inevitable feature of many fully recognised fields of the natural sciences. Astronomy and the theory of evolution are canonical examples, and there are several others. Harold Jeffreys, who was a leading astronomer as well as philosopher of science and probability, argued that this distinction between experiment and observation was not fundamental to the empirical method of science, but was merely 'a difference in technique, and not of principle'.[190]

The fundamental point is not whether the observation is controlled or passive, but whether observations are available in a form that allows for the impact of different factors to be isolated and measured so as to permit the identification of empirical generalisations and the empirical testing of a theory's predictions. This is indisputably the case in fields such as astronomy. Is it also true of fields such as sociology or anthropology? The answer is, perhaps unsurprisingly, not so clear-cut. The opportunity to observe the impact of well-determined changes in single factors on single objects – the equivalent of observing the impact of the changing proximity of Uranus on the orbital path of Neptune - tends not to arise in the social sciences. Observation in the social sciences will more likely involve many observations of an effect in circumstances in which many factors are changing. A key reason for the need for many observations arises because the object of study in social sciences is often an aggregate or a category of objects that are only homogeneous in particular limited ways,

---

[187] See Turnbull (2017), p. 9-15 for further discussion of Graunt's mortality analysis.
[188] Mill (1844). Note this essay was first printed in 1836 but is now mainly available in a collection of essays published in 1844.
[189] Jeffreys (1937), p. 38; Friedman (1953); Nagel (1979), p. 452.
[190] Jeffreys (1937), p. 208.

for example, male actuaries in their (very) early forties, rather than a specific individual thing, i.e. Craig).

Providing the size of all the relevant factors can be measured across all the observations, this data *may* still provide a robust basis for relationships between phenomena to be identified, and hence for the quantification of an empirical generalisation, or the testing of a scientific theory. However, it may not generally be possible to observe and quantify all the relevant factors that may affect human behaviour. For example, the behaviour of individuals may be argued to inevitably depend upon their individual histories and personal experiences. The totality of potentially relevant historical experiences may be beyond observation and measurement.

Moreover, over the course of a historical period, different factors may vary in relevance. As noted in Chapter 3.1, the 'non-stationarity' of human society is a fundamental difficulty for the positivistic approach to social science. Science is based on the basic idea of identifying repeating phenomena. If the causes of the empirical phenomena of social sciences are in a continuous state of flux, there seems little hope of developing theories with empirical predictive strength – the essence of positivist scientific method.

The rigorous quantification of the effect of individual factors across (random) samples of multiple observations is the natural domain of statistical analysis, and it will come as no surprise to find that statistics plays a significant role in positivist social science. Of course, the standard procedures of statistical inference must make some assumptions about the properties of the sample observations. And whilst controlled laboratory experiments may not be logistically or perhaps even ethically feasible in many areas of social science, it may be possible to implement a form of empirical study that can generate a sample of observations that has the characteristics required for rigorous statistical inference (that is, that the sample data is independent and identically distributed, at least from an objectivist perspective). The standard form for this type of empirical study is the *Randomised Controlled Trial* (RCT). In recent decades, the RCT has become a 'gold standard' methodology for empirical observation in a number of major social sciences such as sociology and some fields of economics (indeed, it has been claimed to be at the root of a 'credibility revolution' in applied microeconomics[191]).

Randomised Controlled Trials (RCTs)

Randomised Controlled Trials (RCTs) have become one of the most significant empirical techniques of the social sciences over the last thirty years or so. The statistical theory that underpins the RCT methodology is provided in the statistical inferential techniques that were developed in the early 20th century, most notably the seminal work of Ronald Fisher in the 1920s and 1930s. Fisher was probably the most influential statistician of the 20th century. Books of his such as *The Design of Experiments* had an enormous impact on how statistical methods such as hypothesis testing were applied by researchers across a wide array of disciplines. Fisher's own work was not particularly concerned with topics of social sciences – his focus tended to be on areas of applied natural science such as the agricultural trials of alternative farming methods. But the use of hypothesis testing gradually spread widely over the course of the 20th century, first to fields such as medical and drug trials, and then quite broadly into areas of social scientific study.

The basic RCT technique used in social science involves first identifying a pool of a large number of experimental subjects that have some common forms of homogeneity (for example, 7 year-old Texan school children) and then *randomly* allocating each of these subjects into two groups: the

[191] Angrist and Pischke (2010)

treatment group, which will be exposed to the potential causal factor under investigation (for example, smaller class sizes); and a *control group*, which will not (and, where relevant and possible, the subjects will not be aware of which group they have been allocated to). The randomisation of the allocation seeks to ensure that both groups have statistically similar distributions of any other characteristics that may be potentially relevant to the impact of the causal effect on the subject population (for example, gender, race, IQ, emotional well-being, and unknown casual factors that have not been identified).

The random nature of this allocation process is a fundamental characteristic of RCTs. If all the other potentially relevant factors were known and could be measured, the randomization process would be superfluous and inefficient: in theory, the experimental subjects could be allocated between the two groups in a way that, by design, resulted in the same weights in these various characteristics being present in the treatment group and control group. However, this would require a full tabulation of all potentially relevant characteristics of the subjects and an assessment of their values for each subject. The social scientist simply may not know what factors matter or have the facility to measure them. The randomisation process avoids having to answer these questions. But it inevitably results in some statistical noise in the characteristics of the two groups (that is, the weights of each group will not be equal as some random sampling error will be present). As is the norm in statistical testing, the basic strategy for the reduction of the impact of statistical noise is to make the sample size large enough to give the tests adequate statistical powerful.

If the effect under investigation (for example, student performance in a school test) is observed to occur in the treatment group in a statistically significantly different way to the control group, we then may conclude that there is an empirical (and possibly causal) link between the factor and the effect. Note this effect will almost always be probabilistic in nature (that is, the probability of passing the test may or may not be observed to change, but it usually won't be observed to be 0 or 1 in either the treatment group or the control group). The effect under consideration will rarely produce the same impact on all subjects. And the RCT will not explain why the effect has been observed in some subjects and not others within both the treatment group and the control group.

RCTs are often used to attempt to rigorously measure the impact of some piloted policy intervention – for example, to measure the impact of reduced class sizes on measurable academic outcomes for school children. In this broader setting, the RCT is closely associated with another three-letter acronym of the social sciences and social policy – Evidence-Based Policy (EBP)[192]. EBP reflects a demand from the ultimate intended consumers of social science (policymakers and politicians) for objective, reliable and 'scientific' insight that is not tainted by an over-reliance on subjective expert professional judgement. The EBP field originated in evidence-based medicine which sought to 'de-emphasise intuition, unsystematic clinical experience' and replace it with 'the conscientious, explicit and judicious use of current best evidence'[193].

So, the RCT can be seen as a means of objectively and rigorously gathering empirical evidence in fields where the conventional laboratory experimental techniques of the natural sciences are not feasible. The gathering of empirical evidence is an essential component of the execution of the scientific method. How much can the RCT technique deliver in reliable empirical observation? Is it enough to provide the data for empirical generalisations and the testing of scientific theories of the social sciences?

---

[192] Cartwright and Montuschi (2014), Chapter 3, p.50.
[193] See Reiss (2013), p. 197.

Before discussing possible answers to these questions, we should first recognise that there are many potentially interesting empirical relationships in the social sciences that simply cannot be directly assessed through the RCT methodology. This may be because the empirical factors and effects of interest are too big in scale to be created and controlled in an RCT environment. For example, the impact of central bank interest rate policy on the inflation rate cannot be directly studied using a RCT. There may also be ethical considerations that mean that the methodologically preferable RCT is not appropriate or even legal – for example, implementing an RCT to assess the impact on human health of abusive use of a controlled substance is fraught with ethical and legal difficulties. And it would be practically difficult (or at least time-consuming) to use an RCT to assess empirical relationships whose duration is measured over long time periods such as decades.

We should also recognise that the RCT is an inherently statistical process. It is generally framed as a form of hypothesis test. That is, it aims to answer the question: is there sufficient evidence to reject the null hypothesis that the factor being tested has no effect? As such, the RCT suffers from the basic limitations of any hypothesis test: there are a number of design choices (the null hypothesis and *p* value, for example) that involve some subjectivity; and, related to this, there is the inevitable statistical possibility of false positives and false negatives. When the RCT is defined as the 'gold standard' methodology, it may be easy for the non-statistician to lose sight of the fact that its conclusions may be wrong in a meaningful proportion of cases and that this property is inherent to its design.

The RCT method does not escape from the challenges and limitations that are general and fundamental for the methodology of the social sciences. Chapter 2 discussed how some form of assumption about the uniformity of nature was necessary to allow empirical observation to be used inferentially in science. Intuitively, this assumption does not necessarily travel well when put in the context of the social sciences. Empirical observations of the behaviour of a given society at a given point in time are not necessarily a reliable basis for inferences about the behaviour of another society at another time. So, at the least, it would seem that the generalisations offered by theories of the social sciences and confirmed by RCTs may have to be carefully restricted in their scope of application (in the jargon, they have low-order generality). The history of social sciences and RCTs is replete with examples of where this limitation has become apparent. To take a well-documented case study[194], a large-scale RCT was run in Tennessee in the United States in the mid-nineteen-eighties in which over ten thousand school children were subjects. The purpose of this RCT was to gather empirical evidence on whether reduced class sizes had a positive effect on class performance. The study delivered the intuitive and now newly evidence-based conclusion that smaller class sizes do indeed result in better student performance. Following the publication of the study, a number of other US states then implemented education policies that focused on reducing class sizes. But the effectiveness of the reduced class size policies varied widely across different states, and it sometimes had the opposite effect to the predicted one. This was because the other conditions necessary for the causal effect of the reduced class size were not always present. For example, in some states the policy necessitated a large influx of new teachers that led to a consequent deterioration in the quality and experience of teaching staff.

Confidence in the general applicability of RCTs' conclusions may be increased by obtaining similar results and conclusions consistently across a number of independent RCT studies. But the inference that those results are relevant to a particular new set of circumstances must always rely on a judgment that the conditions are sufficiently similar to the previously tested ones.

[194] Cartwright and Montuschi (2014), p. 55-63.

So, is the complexity of a typical set of social circumstances inevitably too complicated to be sufficiently controlled and replicated to a degree that can support an understanding of causation, explanation and prediction? This has been a long-standing philosophical outlook of some empiricists. For example, several decades before Ronald Fisher's conception of hypothesis testing and the later development of RCTs, John Venn expressed his scepticism about the applicability of statistical approaches to empirical data in the social sciences:

"In all social applications of probability, the unchangeable causes can only be regarded as really unchangeable under many qualifications. We know little or nothing of them directly; they are often in reality numerous, indeterminate, and fluctuating; and it is only under the guarantee of stringent restrictions of time and place that we can with any safety attribute to them sufficient fixity to justify our theory…as a further consequence, the appeal to statistics has to be made with the caution in mind that we shall do mischief rather than good if we go on collecting too many of them."[195]

However, Comte, writing a couple of decades before Venn, expressed a firm belief in the 'unchangeable order of the world' and the 'invariable laws' that govern social phenomena. He recognised that positivism was predicated on the assumption that 'all events whatever, the events of our own personal and social life included, are always subject to natural relations of sequence and similitude'[196]. He also argued this was the case despite the continual changes in society: 'The extensive modifications of which society admits, go far to keep up the common mistake that social phenomena are not subject to any constant law'[197].

Comte's attempt at squaring the circle of constant laws of social phenomena in the presence of continuously changing societal conditions involved distinguishing between irreducible, simple, abstract classes of laws; and compound or concrete laws. He argued it was only the former that were invariable, and it was only the latter that were affected by changes in society. This is rather similar to the distinction that we noted John Stuart Mill made between fundamental laws of human nature and derivative empirical laws of societal behaviour in his discussions of historical explanation (see Chapter 3.1). However, Mill took these fundamental laws to be *a priori* or axiomatic, whereas Comte argued they could only be discovered 'by a series of direct inductions from history'[198]. And, of course, the difficulty with using this line of reasoning as a basis for contemporary social science methodology is that most contemporary positivist activity in social science is focused on phenomena that would naturally change with societal changes rather than on these laws of the fundamental, abstract, invariable kind that Comte suggested exist.

Objectivity and social science

It is sometimes argued that social scientists' own social and cultural outlook, ethical values and political views can result in a bias or prejudice that impacts on their scientific output. This would imply that the output of the social sciences may lack objectivity; that two social scientists, when provided with the same (inevitably limited) empirical evidence, may find support for whatever scientific theory better fits their social, political or cultural preference.

There are two essential phases of the scientific method in which a lack of objectivity may influence the actions of the social scientist: in the formulation of a scientific theory; and in its testing. As discussed in Chapter 2, there are few 'rules' around how a scientific theory is formulated. It is

[195] Venn (1888), p. 94.
[196] Comte (1912), p. 10.
[197] Comte (1912), p. 11.
[198] Comte (1912), p. 18.

essentially a creative process and its legitimacy is not directly determined by the extent to which non-scientific factors, including the scientist's normative values, have influenced this creative act.

It is the process of gathering and appraising empirical evidence in the testing of a scientific theory that demands rigorous objectivity. We saw in Chapter 2.4 that some philosophers of science such as Thomas Kuhn have argued that social and personal incentives may create a reluctance on the part of the leaders of scientific communities to reject long-established theories in the face of falsifying empirical evidence. There are at least two reasons why this effect has the potential to be greater in the social sciences than in the natural sciences: the limitations on the gathering of empirical evidence in the social sciences may create more latitude for a diverse set of semi-testable, competing theories to co-exist indefinitely (the absence of a 'crucial experiment'); and secondly, the social, cultural and political views of the social scientist are often of more relevance to their field of study than is the case in the natural sciences. For example, the extent of a scientist's personal preference for libertarianism or communism is unlikely to pre-dispose them towards one particular version of quantum theory over another, but it may have an effect on their scientific appraisal of Keynes' theory of international trade. Of course, these two reasons are interrelated: the potential impact of their personal views on the social scientist's scientific appraisal is only possible because there is no crucial empirical evidence that would otherwise prevent it.

Max Weber, whose views on the methodology of the social sciences will be discussed further below in Chapter 3.3, argued that the impact of the normative views of social scientists on their scientific output was not simply due to prejudice or bias on the part of the scientist. He argued that it was often intrinsically difficult for the content of a social science such as economics or sociology to fully distinguish its positive empirical statements and logical deductions from its (sometimes implicit) normative value judgements (which Weber defined as "practical evaluations regarding the desirability or undesirability of social facts from ethical, cultural or other points of view"[199]).

Weber argued that this distinction, whilst difficult, was also essential as it formed the boundary between social science and philosophy or political policy (or what Weber referred to as social philosophy): logical deduction and empirical facts could determine the most efficient means of attaining a specific set of ends (science), but those ends were inherently normative and based on ethical and cultural (i.e. non-scientific) values. Thus, according to Weber, the subject of economics could never determine the 'right way' for a society to organise its resources, and *it should never try to*. Rather, it should only seek to determine how to most efficiently achieve a given (normative) objective in the presence of some assumed conditions by choosing amongst a set of means which are otherwise identical in all aspects of value judgement.

The Impact of Observation and New Knowledge on Empirical Relationships in the Social Sciences

The very act of observing in the social sciences may influence the result of the observation. In particular, the result of empirical observation and the development of new knowledge that arises from it may impact on future human behaviour in a way that renders the observations *obsolete and invalid for use in subsequent explanation and / or prediction*.

To take a simple example from the world of financial economics, if a finance professor finds a market-beating stock-picking algorithm and publishes it in a journal, we could expect market prices to then adjust such that the algorithm would not perform so well in the future. Macroeconomics provides the most famous example of this kind of 'feedback' effect in the shape of Robert Lucas'

[199] Weber (1949), p. 10.

Nobel-prize winning model, which gave rise to the 'Lucas critique'[200]. The model shows how governments' attempt at controlling the rate of inflation with the objective of managing unemployment levels will alter inflation's causal relationship with the rate of unemployment as employers will interpret the inflation signal differently in such circumstances. By acting on the knowledge of how inflation and unemployment have related to each other in the past, their future relationship may be fundamentally altered.

More generally, there is an inherently reciprocal or feedback relationship that arises in the social sciences between the development of theory and its application. This type of effect has no parallel with the natural sciences – the extent to which civil engineers use Newtonian mechanics to design and build bridges has no bearing on how accurate Newtonian mechanics is as a description of the physical phenomena the engineer is concerned with. But in the social sciences, acting on the insight provided by the scientific output can materially change the behaviour of the social system that has been analysed. This fundamental idea has been recognised for a long time. For example, John Venn, in 1889, wrote:

"The publication of the Nautical Almanac is not supposed to have the slightest effect upon the path of the planets, whereas the publication of any prediction about the conduct of human beings…almost certainly would have some effect."[201]

This feedback effect is another source of non-stationarity in social systems that can invalidate the uniformity of nature assumption that is implicit in the scientific method.

Assumptions about human behaviour

The social sciences are fundamentally concerned with understanding the behaviour of interacting people. A scientific theory that seeks to explain some feature of human behaviour must make some axiomatic assumptions about the basic motivations or dispositions of people – these will form the premises from which explanations and predictions of behaviour can be logically deduced. An explanation of empirical facts arising from aggregate human behaviour must be derived from some sort of more fundamental model of how individual humans behave. This inevitably must involve some degree of abstraction. Fritz Matchup puts the point succinctly: 'explanation in the social sciences regularly requires the interpretation of phenomena in terms of idealized motivations of the idealized persons whose idealised actions bring forth the phenomena under investigation'[202].

Whilst the axioms of Euclidean geometry may be regarded as sufficiently self-evident for most real-life applications of geometry, the axioms of human behaviour that form the backbone of, say, microeconomic theory (such as the assumptions that individuals can order by preference all consumption choices at all times and will behave rationally and selfishly in light of this knowledge), may not correspond so obviously with the real-world. The drivers of actual human behaviour will, at least in part, be driven by desires and motivations that are normative, that reflect the particular ethical beliefs and values of individual people. Human behaviour is also a function of people's factual beliefs, which may or may not be accurate representations of the facts at a given point in time.

These drivers of behaviour may be neither self-evident nor empirically discoverable. Beliefs are difficult to observe (they are private to the individual), subjective (they are specific to the individual) and they may not remain constant for long (society and the world is continuously changing, and generally in unpredictable and often in unintended ways). Moreover, it might be argued that the

---

[200] Lucas (1976)
[201] Venn (1889), p. 576.
[202] Machlup (1978), p. 353.

existence of human free will is incompatible with universal generalisations about human behaviour: groups of human beings are not homogenous collectives; rather, each human is profoundly unique, each some function of their own personal set of historical experiences that is not accessible to observation. Some element of their behaviour must therefore be unpredictable and idiosyncratic and the explanation of their actions similarly so. Whilst this non-homogeneity argument may logically remove the possibility of the existence of deterministic general laws of human behaviour, it does not necessarily rule out forms of probabilistic or partial explanation as discussed in Chapter 2.3 (though this, in turn, does not necessarily imply that the quantification of the relationship through the use of positive techniques will be more illuminating or reliable than a 'common sense' understanding of the phenomena).

Given social scientific theories' need for axiomatic assumptions about human behaviour, the above difficulties may represent a serious threat to the prospects of success for explanatory theories in the social sciences. It raises the spectre of tautology: that the only explanation for the specific action of a particular person is to make the empty observation that the action must have been the result of the beliefs (factual and / or moral), motives and intentions of that individual at that point in time.

Different views exist on what methodological strategy that is likely to offer the best hope of success in overcoming these difficulties. The most common approach used in the deductive theories of social science is some form of *methodological individualism*. Methodological individualism generally assumes humans are a homogeneous collection of individuals, and the approach focuses on assumptions for the behaviour of these individuals and less on the impact of institutional factors. The standard implementation of methodological individualism in the social sciences, and especially in economics, assumes universally rational behaviour in how all humans form positive factual beliefs and assumes they always act rationally according to these beliefs. This is how the vast bulk of microeconomic theory has been developed and, as noted above, economics is the only subject area of social sciences that has generated notable output in the form of formal, deductive, explanatory theories.

Of course, it is clear that the assumption that permanent, perfect rationality is present in all individuals can only ever be an approximation to reality[203]. The question of how well it corresponds to reality, and, more importantly, how well economic theory's explanations and predictions correspond to reality, is a discussion which will be more fully discussed in Chapter 4.

Closing Thoughts on Positivism in the Social Sciences

Positivist social science is concerned with the identification of empirically recurring relationships in forms of social behaviour of humans and the development of explanatory theories that generate quantitative predictions and generalisations that are capable of empirical testing. The methodological difficulties discussed above, which arise to varying degrees across the social sciences, do not make positivist social science logically impossible. But they represent a formidable battery of challenges that do not arise in the natural sciences (at least not to anywhere near the same degree).

Some have argued that these issues collectively mean that the positivist approach to social science is not likely to be widely successful. In this case, we are left with two possibilities – abandon the ambition of deeming much of social study capable of meeting the criteria of scientific study; or apply

---

[203] The rational behaviour of a given individual is also a function of how others behave, and in particular whether other individuals also behave rationally or not. This leads to the theory of games as first developed by John Nash and John Von Neumann in the 1940s. See Von Neumann and Morgenstern (1947).

a radically different form of methodology which has no recognisable relation to the scientific method of the natural sciences. There is a school of social science that first emerged in the late nineteenth century that advocates just such a radical methodological alternative. It is called *interpretivism* and some of its key features are outlined below in Chapter 3.3.

3.3 Interpretivism

Positivism has long had a methodological rival in the social sciences in the form of what is often referred to as *interpretivism*. We say methodological 'rival' because philosophers and social scientists, like actuaries, enjoy heated debate amongst themselves. But we may more productively view these distinct perspectives as complementary approaches that offer different tools for different jobs. This, we found earlier, could also be said of the objective and subjective definitions of probability; and on the Bayesian and objectivist / falsificationist perspectives on scientific knowledge; and, at the risk of spoiling the conclusion, so too will it be our conclusion for positivism and interpretivism. But this still leaves much to explore and determine: what is interpretivism and how does it differ from positivism? What conditions are necessary for each of the approaches to have a reasonable expectation of success? What 'type' of knowledge can each approach deliver, and what limitations are inevitably placed upon it? These are the types of questions we shall attempt to explore in this section.

The development of interpretivism as a methodology of social science is usually most associated with the German historian / economist / philosopher / sociologist / lawyer Max Weber, who wrote extensively about it at the start of the twentieth century. But the origins of interpretivism can be located in the writings of an earlier generation of Kant-inspired German philosophers, who first wrote of the idea of *Verstehen* (literally, 'to understand'). The German Historicist school of economists, which first emerged in the 1840s and reached its zenith under Schmoller in the 1870s, provides the best early example of the adoption of interpretivist methodology in social science.

So, what is interpretivism? Let's begin by summarising a couple of the particularly notable strands of Weber's philosophical outlook on social science methodology that help to characterise interpretivism as it is understood today.

Weber's Antipositivism

Weber's philosophical outlook was unambiguously *antipositivist*. That is, Weber believed the positive scientific method generally could not deliver reliable scientific output when applied to the social sciences.

Weber believed that human society was infinitely complex and in a constant state of flux (technological, cultural and political). As such, it would be impossible to find scientific laws that could explain and predict interactive human behaviour over different periods of time or across different societies (in the words of Chapter 1, this constant flux made the assumption of a uniformity of nature invalid in the domain of the social sciences). This meant all knowledge in social science must be transitory – even if our understanding of some form of today's social human behaviour is correct, that does not necessarily mean it will still be similarly correct tomorrow.

Naturally, this philosophical outlook made Weber highly sceptical of the value of a methodology for social science that was based on the empirical observation of mass social phenomena and the application of the scientific method of the natural sciences. To Weber, there were no mass social phenomena – rather, every event was unique and should be studied on those terms. His epistemological outlook was robustly sceptical: “Fundamental doubt is the father of knowledge”[204]. He also believed that social scientists should have the ‘intellectual integrity’ to be realistic and honest about these inevitable limitations that would arise in their fields of study.

Whilst Weber was doubtless the most influential early-twentieth century thinker to reject the application of positivism in social science, he had some diverse sympathisers. A notable example is Frank Knight, the eminent American economist, who viewed positivist social scientists as ‘simply bad metaphysicians’[205]. According to Knight, “the fetish of ‘scientific method’ in the study of society is one of the two most pernicious forms of romantic folly that are current amongst the educated”[206]. This contrasts sharply with the logical positivists’ belief in a ‘unity of method’ (the idea that a single scientific method should be applicable for all forms of empirical investigation).

Friedrich Hayek was another antipositivist economist from the first rank of twentieth century social science scholars. Hayek was an Anglo-Austrian economist and liberal political writer of significant influence during the second half of the twentieth century. He was awarded the Nobel Prize in economics in 1974. His philosophy had roots in the Austrian school of economics, which could be said to be philosophically committed to a deep scepticism of positivism in economics and the social sciences more generally. In his Nobel Memorial Lecture, he said of positivism in social science:

“It is decidedly unscientific in the true sense of the word, since it involves a mechanical and uncritical application of habits of thought to fields different from those in which they have been formed.”[207]

He argued that the use of positive techniques in social science would result in a ‘pretence of knowledge’ that would result in policy mistakes that could cause much long-term harm to society. His arguments regarding the severe limitations of positivism in social sciences will, by now, be fairly familiar: the world is intrinsically non-stationary and highly complex; the ‘true’ drivers of social phenomena are highly numerous, interconnected and are often not observable; and constraining our theories to only consider variables that are easily measurable and observable will arbitrarily limit the power of the theories; explanations that offer a closer resemblance to the truth will be rejected because of a lack of quantitative evidence to support them.

Weber’s (and Knight and Hayek’s) critique of positivism is quite consistent with the well-discussed methodological limitations of positivism discussed in Chapter 3.2 (and the limitations of the application of the covering law model of scientific explanation to historical explanation as discussed in Chapter 3.1). A positivist’s response to Weber’s criticism might note that most of Weber’s work on this subject was produced between 1903 and 1917. Since then, statistical inferential techniques such as hypothesis testing and quantitative empirical methods such as Randomised Controlled Trials (RCTs) have emerged that have revolutionised what positivist methods can potentially deliver (leading some to herald the arrival of a ‘credibility revolution’ in some fields of empirical social science). But much of Weber’s critique is more fundamental and is essentially timeless – it is based on a view of the intrinsic nature of human society. And whilst empirical methods such as RCTs have improved the rigour of observation in social sciences, ultimately they can only deliver empirical descriptions of tenuous durability, and not explanatory theory (which, over one hundred years after Weber’s writings, tends to remain conspicuously absent in the social sciences, especially outside of economics, where its empirical success is contentious, as will be discussed further in Chapter 4).

[204] Weber (1949), p. 7.
[205] Knight (1956), p. 152.
[206] Knight (1956), p. 261.
[207] Hayek (1975), p. 30.

The Interpretivist Alternative

The second key strand of Weber's philosophical outlook is the view that antipositivism does *not* imply that it is impossible to generate scientific knowledge in the social sciences. Weber maintained it was possible to develop an alternative methodology of social science that was capable of "analytically ordering empirical reality in a manner which lays claim to validity as empirical truth"[208]. But instead of being built on the empirical observation of mass phenomena, Weber's alternative was based on the detailed examination of individual, unique and specific 'concrete' cases. Indeed, to Weber it was the *uniqueness* of significant events that was the fundamental source of their interest: "the specific meaning which a phenomenon has for us is naturally *not* to be found in those relationships which it shares with many other phenomena"[209].

This focus on the detailed study of unique occurrences may at first glance appear more aligned to history and historical explanation than the scientific activity of forming empirically testable generalities. Like the historian, Weber was more interested in the explanation of unique, individual historical events or episodes than in the development of general laws for mass phenomena. But above we saw that attempts (of positivist philosophers of science) at historical explanation have generally used generalisations to explain why a specific action occurred (for example, Hempel's covering law model of historical explanation). Weber's belief in the infinite complexity and continuously changing nature of reality meant that general causal laws of the type developed in the natural sciences would be useless for causal explanation in social science. So, Weber's approach to historical explanation was different and was based on an unequivocal rejection of Comte and Hempel's positivist philosophies. The explanation of these individual events could not be delivered by general laws but by identifying "concrete causal relationships" in unique events. And Weber believed an analysis of such relationships could provide insight and explanation into the behaviours that occur in other unique events. But how?

Weber believed an understanding of human behaviour could be obtained through the scientist's application of his or her human sense of empathy: 'the social scientist…is able to project himself by sympathetic imagination into the phenomena he is attempting to understand.'[210] That is, the social scientist is able to place him or herself in the shoes of another, and imagine how it feels and how they would be compelled to act in such circumstances. This is the essence of *Verstehen*. We noted in Chapter 3.1 that Weber viewed the untangling of empirical facts from normative views in social science as intrinsically difficult. In this interpretative method of empathetic understanding, the social scientist would attempt to cast aside his or her own normative views and replace them with those of the object of study. Weber maintained that this interpretative methodology was scientific, empirical and capable of delivering 'causal analysis which attempts to establish the really decisive motives of human actions'[211].

So, interpretivism seeks a meaningful form of explanation of human behaviour through a qualitative, empathetic analysis of the behaviours, values, culture, language, tradition, communities, social structures and so on that have been empirically observed in specific, unique and significant cases, from the perspective of the human participants. Weber's objective for the output of an interpretative analysis was "the sure imputation of individual concrete events occurring in historical reality to concrete, historically given causes through the study of precise empirical data which have

[208] Weber (1949), p. 58.
[209] Weber (1949), p. 76-77.
[210] Nagel (1979), p.484.
[211] Weber (1949), p. 14.

been selected from specific points of view"[212] in order to "understand on the one hand the relationships and the cultural significance of individual events in their contemporary manifestations and on the other the causes of their being historically so and not otherwise."[213]

The Ideal Type

At the risk of making the classic positivistic mistake of unreliably over-generalising about human behaviour, it is likely that the typical Anglo-Saxon actuary may find the concept of *Verstehen* a little bit 'fluffy'. Weber also advocated the use of a methodological tool that was more explicitly logical in its analytical workings. This was the concept of the '*ideal type*'.

The ideal type is an abstraction – it is a mental construction that can be considered as a simplified, isolated or magnified version of some empirical object or concept. Importantly, the 'realism' of the ideal type is not the primary metric of its quality. Rather, its measure of usefulness is found in its ability to provide insight into the explanation of the interrelationships between some phenomena of interest. Weber was not the first to identify the use of this sort of device in the methodology of the social sciences. An earlier example of such an idea can be found in the work of another German economic thinker, Johann Heinrich von Thunen, in his famous economic model of agricultural land, as described in his 1826 book *The Isolated State*. The phrase has historically been used by various philosophers and methodologists since the nineteenth century to the present day. Here we will focus mainly on what 'ideal type' meant to Weber.

Weber's ideal type was not intended as a direct representation of empirical reality or as some form of 'average'. Rather, the ideal type was a simplified model of reality that could act as a form of reference point or benchmark where the behaviour implied by the ideal type could be compared with empirical reality to aid its understanding. In this sense, it is deliberately counterfactual: "this construct in itself is like a utopia which has been arrived at by the accentuation of certain elements of reality. Its relationship to the empirical data consists solely in the fact that where market-conditioned relationships of the type referred to by the abstract construct are discovered or suspected to exist in reality to some extent, we can make the characteristic features of this relationship pragmatically clear and understandable by reference to an ideal type."[214]

So far, this notion of an ideal type may sound somewhat similar to the concepts used in a hypothetico-deductive positive scientific model. All scientific models involve a simplification of reality that generates deductive consequences. Indeed, some methodological writings such as those of the nineteenth century Austrian economist Carl Menger argued that the idealisations used in the theories of the natural sciences (such as perfect vacuums or frictionless motion) were forms of ideal type.

Whilst the concept of idealisation also plays a key part in scientific explanation in the natural sciences, there is a fundamentally important difference: in the hypothetico-deductive scientific method, the consequences deduced from the idealised concepts are empirically tested to determine whether the model is useful for making generalisations and quantitative predictions about the behaviour of other phenomena. But this was not the way Weber envisaged the ideal type being used. To Weber, "the ideal type is an attempt to analyse historically unique configurations or their individual components by means of genetic concepts"[215]. His ideal type is not intended to be as empirically realistic as possible. Rather, it is intended to illuminate how a specific event or episode is

[212] Weber (1949), p. 69.
[213] Weber (1949), p. 72.
[214] Weber (1949), p. 90.
[215] Weber (1949), p. 93.

unique and different from the model of the ideal type and what the consequences of these differences are: "the construct here is no more than the means for explicitly and validly imputing an historical event to its real causes while eliminating those which on the basis of our present knowledge seem possible."[216]

So, by this reasoning, Marxist theory, for example, could be useful not because it provides reliable empirical predictions (it generally does not), but because it can operate as an ideal type that can provide a better understanding of the actual empirical historical development of society: the process of comparing and identifying the differences between the real-world and the world implied by Marxist theory can provide a better interpretative understanding of how the real world works. This key methodological difference between the use of a model as a form of causal law for prediction and the use of a model as an ideal type to generate illuminating comparisons is fundamentally what differentiates interpretivism from positivism[217].

The essential form of the logical output of the ideal type is a comparison of what happens according to the model of the ideal type and what happens (or happened) in a unique example of empirical reality. The differences elucidated by this comparison can be explained by the differences between the model of the ideal type and the real-world that produced the empirical example. But there is a limit to how far this form of explanation can logically take us: by Weber's own description of reality there must be infinitely many differences between the model of the ideal type and the infinitely complex real world. How do we know which of these differences are the important ones that really 'matter' in explaining the differences that have been identified between the deductive consequences of the ideal type model and the observations of empirical reality? The ideal type analysis is not intended to provide a logically complete answer to this question – it is not a deductive solution, but an aid to a form of inductive reasoning about the behaviour of complex empirical phenomena.

Closing Thoughts on Interpretivism

Beyond the limitations of the logic of the ideal type, there are broader difficulties with the interpretivist methodology. Clearly, understanding by sympathetic imagination is not a route to evidence open to the natural scientist who is concerned with the behaviour of physical phenomena. Does the human nature of the phenomena that is studied in the social sciences really provide a novel source of evidence that arises from one human's ability to imagine himself in the shoes of another?

The argument of 'understanding by sympathetic imagination' raises concerns about objectivity. This understanding is not empirical, at least not in a direct, external and verifiable sense. Some might argue that it is the essence of science that its observations are public and repeatable, and its testing

---

[216] Weber (1949), p. 102.

[217] It was noted above that Friedrich Hayek was another prominent antipositivist. Hayek's methodological solution was distinct to the interpretivist's use of ideal types and may be worth mentioning in comparison. Although antipositivist, Hayek still advocated the development of theories of social science that had direct empirical content. But he believed these theories could only be capable of saying significantly less, and with much less precision, than was claimed by positivist methodology. Nonetheless, he argued that such theories would still have meaningful empirical content, and should still be considered as *bone fide* science, because they would pass Popper's test of falsifiability: although the theories would be less precise about what *would* happen, they would still imply some outcomes *that could not happen*, hence making the theory technically falsifiable. Such a perspective is notable as another alternative to positivism. It is not contradictory to interpretivism – an interpretative ideal type and a falsifiable Hayek theory could co-exist and even be mutually supportive – but it is a logically distinct resolution to the problems of positivism in the social sciences.

is interpersonal (i.e. the same tests and results can be obtained by anyone). But interpretative understanding is an internal knowledge that is developed from the individual empathetic reflections of the social scientist. It would seem natural that the scientist's sense of human empathy would be influenced by their own life experiences. This suggests an interpretative analysis may have at least some element of subjectivity. And it would appear logical to infer that this skill of empathy would be stronger when analysing societies that are closer in form and time to the scientist's own experiences, which implies a potential relativity. How can we determine if one person's interpretative account is better or more valid than another's?

Where does this brief overview of the ideas of interpretivism leave us? It seems clear that intrepretivism is primarily interested in the explanation of unique events, whilst positivism is primarily engaged in attempts at rigorous observation and measurement of mass empirical phenomena, potentially with a view to explaining and predicting them. The empiricism of intepretivism is therefore far removed from the empiricism of the positive scientific method. This does not mean it is not meaningful, and it is natural that some knowledge in the social sciences must take this form, especially in circumstances where the conditions are not ripe for successful positive science. Even Hume, that most sceptical of empiricists, wrote economics essays full of causal explanations that were justified by reference to specific historical experiences[218].

Interpretivist methods may provide powerful, rational and analytical insights into why humans behave in certain ways in specific circumstances. It may be viewed as an impartial, systematic method of the social sciences. It may provide explanations that can be considered intellectually robust, highly valuable and entirely valid. It might even be argued that interpretivism is capable of generating a form of probable knowledge. But interpretivism is generally incapable of delivering explanation in a form that is capable of empirical testing, confirmation or falsification. It therefore does not meet some of the essential, basic criteria of science. Indeed, it could be argued that the interpretivist approach is based on the premise that science is not possible in the study of the social behaviour of humans. This is an entirely reasonable philosophical view (though it wasn't how Weber characterised it). But it would be terminologically simpler if the output of interpretivist analysis was called something other than scientific knowledge. So, what might it be called? Interpretivist thinkers such as Frank Knight offered a range of terms for this form of knowledge: common sense, informed judgment, human insight. Perhaps interpretative knowledge is a natural general term. Whatever we call it, it is something qualitatively different to positivist scientific knowledge: that is, knowledge that has been acquired through a successful application of the scientific method as outlined in Chapter 2. Both are forms of empirical knowledge and understanding that have been acquired from experience using methods beyond deductive reasoning. But the differences are meaningful.

[218] See Hoover (2001), Chapter 1.1 for a fuller discussion of Hume and causation in macroeconomics.

4. On the methodology of economics

This chapter builds on Chapter 3's discussion of methodology in the social sciences in general by considering how those topics apply to the specific social science of economic. Economics is the social science concerned with how human society produces, distributes and consumes its wealth of scarce resources. It is often claimed to be the most advanced of the social sciences. It is generally recognised as the social science that has made the greatest progress in specifying deductive theoretical systems that aim to explain social phenomena in a manner consistent with the positive scientific method of the natural sciences. As a result, many of the topics of Chapter 2 on the application of the scientific method to the natural sciences will be of relevance in this chapter as we consider the methodological issues that can arise in economics. At the same time, economics is unambiguously a discipline of the social sciences, not the natural sciences, and the topics of Chapter 3 will therefore be very relevant to economic methodology.

Given the relatively advanced nature of economics as a social science, and its relevance to major fields of actuarial theory and practice, this chapter has been specifically dedicated to the methodology of economics. Again, we shall focus on the aspects of methodological doctrine and dispute that are likely to be of greatest interest to actuarial science. In particular, we have focused on topics in which the economic content is germane to the content of actuarial science (for example, econometrics and option pricing); and on topics where the form of methodological question has implications for broad areas of actuarial methodology (most obviously, in the long-term empirical projection of social and economic phenomena).

A brief potted history of the development of economics as a field of study is provided in an appendix. This historical background may be helpful in gaining a perspective on the methodological developments and challenges that have accompanied the subject as its techniques and schools of thought have evolved (sometimes quite radically) over the subject's history.

4.1 Economics as a positive social science

The deductive structure of its theories, its highly quantitative content, its ambition to predict and explain observable empirical phenomena, all suggest economics has more than a passing methodological resemblance to natural science. And whilst various schools of economic thought have rejected the notion of economics as a positive science capable of performing accurate empirical prediction, it seems clear that mainstream orthodox economics is considered by its practitioners as a positive scientific discipline with methods capable of producing scientific laws. Since the time of Ricardo, the core output of mainstream economics has been the development of abstract deductive theoretical systems that seek to deliver general explanation and prediction based on 'a formulation of what holds true under given conditions'[219] rather than mere statistical generalisation or specific historical analysis.

Like any deductive system, economic theories will explicitly (and implicitly) specify some premises, postulates and assumptions from which implied consequences for the phenomena of interest will be deduced. Since the Marginal Revolution and the emergence of neoclassical economics[220], the articulation of most important economic theories and their deduced implications has tended to be undertaken in a formal logico-mathematical setting (the work of Keynes was perhaps the greatest departure from this; but decades of Keynesian economists followed him to pick up the task of expressing his rhetoric more formally).

[219] Fisher (1911)

[220] See the Appendix for some further description of the historical development of economics.

As the only social science to make serious progress in the development of deductive theoretical systems, economics has encountered some unique methodological challenges. Some of the most notable and enduring of these are discussed below.

Deductive Theoretical Systems in Economics – Behavioural Postulates and Factual Assumptions

As is the case for any deductive scientific theory, in developing an economic theory the economist will aim to begin with the minimum and most elementary of premises from which will be derived a set of predictive implications that are as wide and as accurate as possible for some phenomena of interest. An economic theory will typically specify a set of basic postulates that describe the assumed nature of its theoretical world and the economic objects the theory is concerned with (for example, individuals, markets, firms, or the macro economy at large). As economics is an intrinsically social science, this will necessarily include some explicit or implicit assumptions about the behaviour of human individuals.

Most orthodox economic theory uses a form of methodological individualism in which behavioural postulates specify assumed behaviour for individuals *en masse* – that is, individuals are considered as an aggregated pool of many generalised anonymous actors. However, it is also possible for some economic theories, especially those in macroeconomics, to either explicitly or implicitly make assumptions about the behaviour of institutional or political actors. These could be interpreted as postulates about the behaviour of a much narrower and more specific set of individuals (for example, a central bank governor), and this latter type of premise may be viewed as being less reliable, self-evident or valid as a basis for theory[221].

Standard neoclassical economic theory assumes that individuals know what ends they wish to achieve (which may be expressed as the ability to rank their consumption preferences). Economic theory takes the ends as a given set of preferences, it does not study what ends individuals *ought* to have. It assumes individuals have control or choice over the means of going about achieving those ends and behave in knowledgeable, rational ways that are solely motivated by an objective of using those means to realise the ends to the maximum extent. This can be expressed as an objective of maximising the 'utility' of the consumption of their wealth.

A number of other basic postulates of individual behaviour beyond perfect knowledge, a preference for more rather than less wealth and rationality characterise orthodox economic theory. These include an assumption of diminishing marginal utility of wealth (or, equivalently, diminishing marginal rates of substitution for all commodities and services); and an assumption that all commodities and services are ultimately scarce. Assumptions about the behaviour of firms as well as individuals will also be present in orthodox microeconomics. In particular, it may be assumed that firms seek to maximise profits (or shareholder value) and that firms experience a law of diminishing returns as they increase the quantity of any of the inputs into a production process.

Since nineteenth century classical political economy, the characterisation of humans as knowledgeable, rational and solely motivated by wealth is sometimes referred to as Economic Man or *homo oeconomicus*. Economic Man is motivated by wealth and maximising the utility obtained from that wealth. For the avoidance of doubt, Economic Man is not influenced by anything else such as 'the love of a certain country, or a certain locality, inertia, habit, the desire for personal esteem, the love of independence or power, a preference for country life, class prejudice, obstinancy and

[221] See Machlup (1978), Chapter 2 for fuller discussion of this point.

feelings of personal ill-will, public spirt, sympathy, public opinions, the feeling of duty with regard to accepted obligations, current notions as to what is just and fair'[222].

The partial and isolated set of motivations that define Economic Man has been responsible for an enormous amount of philosophical and methodological debate, particularly about whether Economic Man is intended as a realistic, factual description of the economic decision-making of people, or whether he should be considered as a heuristic fiction or ideal type in the Weberian sense. Those who are sceptical of the notion that economics can function as a positive science have not found it difficult to highlight the limitations of Economic Man. This point will be discussed further below. But, for our present purpose, it can be immediately noted that the above brief discussion highlights two key boundaries that economists have conventionally set when specifying the scope of their discipline: how individuals determine and order their preferences (their ends rather than the means) is considered a topic for others (perhaps other social sciences); economics is only concerned with human actions that are motivated by a *specific sub-set* of human incentives (those that arise from a desire to maximise the utility they experience from consumption of their wealth).

This latter point on the *partial* nature of economics has been consistently applied by (classical and neoclassical) economists since the start of the nineteenth century (if not, before). As was noted earlier, the late nineteenth century and its following decades saw a proliferation of different schools of economists, some of which, such as the American Institutionalists and German Historicists, had philosophical and methodological foundations that were strongly opposed to a positivist methodology. They emphasised studying the fullness of realistic behaviour as found in historical episodes, emphasising that different times have had profoundly different societies, cultures and institutions. These approaches therefore tended to embrace a wider set of social sciences in their study, such as sociology and history, and were set against the partial perspective of the classical and neoclassical schools. The debates between the 'classical' English school and these new schools were often tainted by highly emotive and bitter language. Somewhat similarly, but from a quite different philosophical perspective to the historicists, Auguste Comte also argued that economics should be subsumed into a more general social science that takes account of all aspects of social life.

But the mainstream of economics has consistently advocated and implemented a partial scope for economics for the last two hundred years. Whilst the content of economic theory may have developed unrecognisably, the focus on the economic implications of a partial sub-set of the possible causes of human action has been a permanent and continuous feature of positive economics. In his essay of 1836*, On the Definition and Method of Political Economy*, J.S. Mill unambiguously positioned economics within the broader field of social sciences as a sub-discipline that only focuses on behaviour motivated by wealth:

"It [economics] does not treat of the whole of man's nature as modified by the social state, nor of the whole conduct of man in society. It is concerned with him solely as a being who desires to possess wealth, and who is capable of judging of the comparative efficacy of means for obtaining that end. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth."[223]

This leads naturally to Mill's definition of economics:

[222] J.N. Keynes (1891), p. 124-6.
[223] Mill (1836), p. 41.

"The science which traces the laws of such of the phenomena of society as arise from the combined operations of mankind for the production of wealth, in so far as those phenomena are not modified by the pursuit of any other object."[224]

As is demonstrated by the caveat clause at the end of his definition, Mill, and indeed the vast majority of economists that have followed since, have recognised that there will inevitably be a 'plurality of causes' at work in the real world beyond those contained in the sub-set of causes that economists have decided is within the domain of their subject. As a result, Mill argued, it would only be possible for economic theories to predict 'causal tendencies' – that is, what would happen if the given causal factors changed and nothing else changed, or, as an economist frequently and succinctly puts it, assuming *ceteris paribus*.

The function of the *ceteris paribus* assumption in economic theories has been an area of some methodological consternation. To understand these concerns, we must recognise that the clause can be used in a couple of different ways: it can be used to say that, if all factors in the system remain unchanged, then a given relationship between two or more variables will exist; or it can be used as a more general all-encompassing condition to state that everything else, both specified and unspecified by the theory, remains unchanged.

When the theory is considered only as a self-standing logical deductive system, the clause can only have meaning in the former sense. But even there the meaning can be ambiguous. For example, the effect of a decrease in the marginal propensity to consume may result in either an increase in investment or a reduction in employment, depending on whether *ceteris paribus* is defined to mean total demand remains unchanged or employment on capital goods remains unchanged. It cannot be both, and so this definition is a fundamental and essential part of the theory. In so far as these are simply different factual assumptions that differentiate competing economic theories, there is nothing logically problematic with that. The difficulty lies is in the ambiguity that can emerge when a given theory's assumptions are only partially defined by use of a vague *ceteris paribus* that does not fully specify the *cetera*.

The meaning in the second sense arises when we attempt to find some correspondence between the abstract world specified in the theory and empirical reality. By construction, the partial abstractions of economic theory leave gaps between theory and reality that must be bridged in seemingly arbitrary ways. As a result, this latter use of the *ceteris paribus* clause has come under particular criticism from some philosophers and methodologists, as it makes it very difficult to determine under exactly what conditions an economic theory's predictions are expected to hold true ('*ceteris paribus* is always a construction rather than an operational concept'[225]).

Deductive Theoretical Systems in Economics – Deducing the Implications of the Postulates

Since the Marginal Revolution of the 1870s, there has been a relentless tendency for economic theory to be expressed in increasingly mathematical language. This can be seen as part of a broader shift towards the use of more formal and rigorous deductive arguments across the entire academy of sciences over the same time period. Even in mathematics itself, Frege in 1879 developed a formal language in an attempt to ensure that deductions from axioms to mathematical theorems were 'free of gaps'[226]. In neoclassical economics, the standard instrument of deduction is a constrained mathematical optimisation. In financial economics, another field of economics that has been

---

[224] Mill (1836), p. 43.
[225] Machlup (1975), p. 195.
[226] Frege (1879), p. iii. Also, See Linnebo (2017) for a fuller discussion of the philosophy of mathematics.

developed within a mathematical framework, the deductive consequences have often been derived directly from the assumption that market prices will not permit arbitrage amongst different financial assets.

The greater clarity of presentation provided by a formal mathematical framework relative to that provided by the rhetorical arguments of the classical political economy era has long been advocated in methodological discussions. Methodologists immediately recognised the benefits of the mathematisation that the Marginal Revolution had brought. In his classic tome of 1891 on economic methodology, John Neville Keynes (father of Maynard) wrote:

"It would be difficult to exaggerate the gain that has resulted from the application of mathematical ideas to the central problems of economic theory."[227]

The advantages in the use of formal mathematical language in deductive reasoning are quite intuitive. It flushes out any logical ambiguities in the rhetoric and this process may even identify logical fallacies. It may bring clarity to the meaning of potentially vague *ceteris paribus* clauses, making assumed mutual dependencies explicit through concise and exact statements. As Irving Fisher, who was perhaps the most influential US economist of the early twentieth century, put it:

"An algebraic statement is usually a good safeguard against loose reasoning; and loose reasoning is chiefly responsible for the suspicion under which economic theories have frequently fallen."[228]

But this perspective was not uniformly shared by all economists of that era. The schools of economics that sat apart from the neoclassical school, most notably the German Historicists, rejected this use of a highly mathematical deductive framework for economic theory. The Historicists' rejection of a mathematical framework for economics, however, can largely be viewed as part of a more fundamental rejection of any form of positivist approach to the subject: instead, their interpretivist philosophy argued that theoretical insights could only be obtained through the study of specific historical events. Nevertheless, scepticism about the extensive use of mathematical techniques in economic theory was not confined to Edwardian-era German interpretivist social scientists. John Maynard Keynes (son of Neville), in his *General Theory*, was quite emphatic on this point and is worth quoting at length:

"The object of our analysis is not to provide a machine, or method of blind manipulation, which will furnish an infallible answer, but to provide ourselves with an organised and orderly method of thinking out particular problems; and, after we have reached a provisional conclusion by isolating the complicating factors one by one, we then have to go back on ourselves and allow, as well as we can, for the probable interactions of the factors amongst themselves. This is the nature of economic thinking. It is a great fault of symbolic pseudo-mathematical methods of formalising a system of economic analysis…that they expressly assume strict independence between the factors involved and lose all their cogency and authority if this hypothesis is disallowed; whereas, in ordinary discourse, where we are not blindly manipulating but know all the time what we are doing and what the words mean, we can keep 'at the back of our heads' the necessary reserves and qualifications and the adjustments which we shall have to make later on, in a way in which we cannot keep complicated partial differentials 'at the back' of several pages of algebra which assume they will all vanish."[229]

---

[227] John Neville Keynes (1891), p. 251.
[228] Fisher (1911), p. 24.
[229] John Maynard Keynes (1936), P. 257.

There are a couple of distinct points that Keynes makes here. One is that the decision to use a mathematical language to express an economic theory will itself constrain the form of assumption that is used: mathematical language is not merely being used to describe the economist's assumptions; rather, the use of mathematical language and the desire for ease of mathematical disposition *constrains the choice of economic assumptions that can be used*. And he is also making another point: economics is not the sort of subject that can provide precise or 'infallible' answers; striving to obtain them is therefore futile. This raises the more fundamental point of whether economics should be considered as a positive science or as something else. For now, we may note that the first sentence of this quotation of Keynes sounds remarkably consistent with the idea of deductive theory as an interpretative ideal type. That is, the theory provides a tool for '*thinking out particular problems*' rather than for generating the accurate empirical predictions that we can regard as the currency of successful positivism. We will return to this point in Chapter 4.2.

The development of economics since the publication of Keynes' *General Theory* in the 1930s has run contrary to his arguments above. Over the second half of the twentieth century, neoclassical economic theory has evolved in an ever more mathematical format. Economics is not unique in this regard. Since the end of the Second World War, an array of phenomena such as the advent of computing power, improvements in data collection and breakthroughs in statistical inferential techniques have resulted in an increasing mathematisation of much of science, finance and other fields of life and study (and this is a trend that we appear to still be in the midst of today).

As a social science seeking forms of deductive theoretical explanation, economics may be one of those disciplines that is especially suited to (or susceptible to, depending on your viewpoint) mathematisation. The barriers to the empirical testing of economic theories (discussed further below) may create an incentive for economists to develop theories that perform well in some other sense, and in a sense that is quite different to the objectives of the scientific method discussed in Chapter 2. The scientific method imposes a strong discipline on the scientist: the theory is only as good as the empirical accuracy of its predictions. Without this disciplining feature, theory-building may be motivated by other incentives. For example, academic economists might find that building aesthetically pleasing theoretical economic systems and deducing interesting properties about such systems (which may provide little insight into economic reality or generate empirical predictions) and presenting these findings to other academic economists is the way to progress in his or her profession.

This lack of 'empirical discipline' may incentivise theoretical complexity. In a well-functioning scientific method, complexity is only tolerated, not encouraged. And it is tolerated only to the extent it can explain more empirical phenomena. In the absence of empirical testability, complexity does not necessarily deliver more realism or a deeper explanation or predictive performance for empirical economic phenomenon, but instead may be rewarded for its own sake. Moreover, economists may not be motivated to tackle the forms of complexity that offer greater realism if such complexity happens to be mathematically intractable ('elegant models of comely shape but without vital organs'[230]). In this scene, the subject may be at risk of devolving from a scientific discipline into a mere exercise in abstract logical puzzle-solving.

Deductive Theoretical Systems in Economics – Empirical testing and the 'a priori' perspective

Chapter 3.2 discussed some of the challenges that arise when attempting to empirically test the predictions of positive theories of social science (most notably, the difficulty in obtaining observations in well-defined conditions). Economics is a social science, and those complications

[230] Karmack (1983), p. 122.

generally apply as readily to economics as they do to other fields of the social sciences. Take a very simple example: it was predicted by the most basic of economic laws that the removal of import tariffs delivered by the repeal of the Corn Laws in 1846 would have the effect of permanently lowering the price of wheat in Britain. But that is not what subsequently happened. Various explanations for this can be found by identifying many 'disturbing causes': the price of gold significantly depreciated in the following years; there were failures of the potato crop; and so on. An economist would naturally argue that the price of wheat was lower than it would have been relative to the counter-factual scenario where the Corn Laws were not repealed. But that is clearly not a directly testable assertion.

There are some special aspects of the subject of economics that may make it even more difficult to test its predictions than in other social sciences. In particular, the above discussion of the basic postulates of economic theory identified a defining trait of economics: it only concerns itself with a sub-set of the causal phenomena that may determine human behaviour in the real world. Given the practical difficulties in isolating these causes in the empirical observation of economic phenomena, and the near impossibility of observation via controlled experiment for most economic phenomena, it is often extremely difficult to empirically test the deduced predictions of economic theories. Infinitely many potentially 'disturbing causes' are always exerting an influence on the empirical data, and many of these influences may be very difficult to even identify, let alone accurately measure.

This has been a recognised difficulty for the methodology of economics since the mid-nineteenth century (if not before). Mill argued that it provided an explanation for why the conclusions of Ricardo's theory were at such odds with the empirical realities of nineteenth century Britain – there was nothing wrong with the logical deductions of Ricardo's theory, it was simply that there were some disturbing causes and circumstances in the empirical outcomes that the theory did not accommodate. Mill's methodological conclusion was that the hypothetico-deductive scientific method could not be implemented in economics (not that it was called that in Mill's time): it was never possible to empirically test the deduced predictions of the theory, as the theory only ever attempted to partially capture the spectrum of causal influences that affect reality. Instead, Mill argued that the intrinsically partial nature of economics meant that the economist must adopt a form of *a priori* method.

The essence of an *a priori* approach is that the degree of belief in the adequacy of a theory is founded on the degree of confidence in its *premises, postulates and assumptions* rather than its *predictions*. If the truth of the postulates are established beyond doubt, and there is no doubt that the logical consequences of the postulates are also correctly deduced, then there is no need to test the theory by considering whether the empirical evidence (which is inevitably contaminated with disturbing causes and circumstances) is or is not consistent with the theory's predictions.

Viewed from this perspective, comparison of empirical observation with an economic theory's predictions can only tell us (usually, *ex post*) about whether the theory is *applicable* in a particular set of realised empirical circumstances. The testing cannot tell us about the *validity* of the theory. This logic also implies that the applicability of a given economic theory is a function of a given society's economic organisation, structure and conditions. As these conditions are continually changing within a given society and can vary widely across different societies at any given point in time, this clearly implies the degree of applicability of a given theory may vary widely.

The *a priori* approach can take many forms. The most extreme version of the *a priori* method in economics has been advocated by the Austrian school (which today, and for most of the last 100 years, is generally regarded as an unorthodox offshoot outside of mainstream economics). The

Austrian school was founded in the 1870s by Carl Menger, but it was the work of Ludwig Von Mises (who we encountered in Chapter 1's discussion of the philosophy of probability theory) in the mid-twentieth century that really clarified its unique methodological perspectives[231]. Austrian economics regards the basic postulates of human behaviour in economics as exact and necessary statements – not merely self-evident, but *known to be true by construction* (essentially because all human action is considered by the Austrians as purposeful action, which they take to be rational by definition). In Kantian language, the postulates are considered to be 'synthetic *a priori*' statements that are beyond empirical refutation. On the appraisal of economic theory, Von Mises argued:

'[Economics'] particular theorems are not open to any verification or falsification on the grounds of experience…the ultimate yardstick of an economic theorem's correctness or incorrectness is *solely reason unaided by experience*.'[232] [italics added]

Whilst the most extreme form of *a priorism* in economics was advocated by the Austrian school which came much later, Mill's writings suggest that a weaker but nonetheless definite form of *a priorism* was deeply embedded in the methodological perspectives of classical political economy.

Many of the early twentieth century's leading economists, spanning a number of schools of the profession, continued to argue that the basic postulates of economic theory should be regarded as self-evident (if partial) truths. Some particularly notable examples include:

- Joseph Schumpeter, in 1914, wrote that the basic assumptions of economics "are so simple and are so strikingly confirmed by everyday experience as well as historical experience that it would be a shame to waste paper printing any special compilations of facts to confirm them"[233].
- Lionel Robbins, one of the leading British economists of his generation, in his highly influential *An Essay on the Nature and Significance of Economic Science* (first published in 1932, with a second edition in 1935), argued that the essential postulates of economics are "simple and indisputable facts of experience" that "have only to be stated to be recognised as obvious"[234].
- Frank Knight, who was a notable sceptic of a positivist approach to economics, argued in 1924 that some of the fundamental postulates of deductive economic theory could be considered as 'facts' – for example, the law of diminishing marginal utility must hold true, otherwise people would expend all of their wealth in the consumption of the first commodity they found[235]. (Although, as we shall read below, Knight disputed the usefulness of Economic Man as a basic postulate.)

Some of these economists argued that when the self-evidence of the postulates of economic theory are considered alongside the severe difficulty in empirically testing the theory's predictions, the only methodological solution was an *a priori* one: empirical data was so noisy and confounded by disturbing causes, it could tell us nothing about the validity of a theory, only whether the circumstances permitted its application.

---

[231] Von Mises (1949)
[232] Von Mises (1949)
[233] Machlup (1978), p. 468
[234] Robbins (1935)
[235] Knight in Tugwell (1924), p. 257.

Robbins' essay stated "the validity of a particular theory is a matter of its logical derivation from the general assumptions which it makes. But its applicability to a given situation depends upon the extent to which its concepts actually reflect the forces operating in that situation."[236]

And Schumpeter wrote in 1949:

"It [empirical statistical evidence] cannot disprove the proposition [of economic theory], because a very real relation may be so overlaid by other influences acting on the statistical material under study as to become entirely lost in the numerical picture…Material exposed to so many disturbances as ours is, does not fulfil the logical requirements of the process of induction."[237]

As may have been noted from some earlier discussions, empiricism was strongly in fashion amongst philosophers of science during the second and third quarters of the twentieth century. It came in a variety of flavours and from a range of advocates - operationalists, logical positivists, Popperian falsificationists and logical empiricists – but they all shared a deeply empiricist outlook. And yet, in economics, a defence of an *a priori* approach can continue to be observed throughout the third quarter of the twentieth century from the leading voices of the profession. In his 1957 essay, *The Construction of Economic Knowledge*[238], T.C. Koopmans argued that the extreme difficulty in empirically testing the predictions of an economic theory meant it was more important to gain confidence in the theory's postulates. Fritz Machlup, a philosopher of the logical empiricist school who has perhaps published more on the philosophy of economics than anyone else, made a very similar point in 1960:

"The difficulty with verification [of economic theory] lies less in the excessive purity of the abstract constructs and models than in the gross impurities in the operational concepts at our disposal. Empirical tests employing inadequate operational concepts are, quite properly, not accorded much weight and, hence, *the theories so tested survive any number of apparently contradictory experience*."[239] [italics added]

Whilst contemporary economists outside of the Austrian school would tend to baulk at the idea that they practiced an *a priori* methodology that was far removed from recognised scientific method, those who have specialised in study of the methodology of economics find it harder to reject the claim that a form of *a priori* method remains at the core of much of economic theory. Daniel Hausman, one of the leading contemporary methodologists of economics, wrote:

"It is not unacceptably dogmatic to refuse to find disconfirmation of economic "laws" in typical failures of their market predictions. When the anomalies are those cast up by largely uncontrollable observation of complicated market phenomena, it may be more rational to pin the blame on some of the many disturbing causes, which are always present."[240]

So, we have been able to trace a consistent commitment to some form of *a priorism* in the methodology of economists from the mid-nineteenth century to contemporary times. This commitment largely rests on the allegedly self-evident nature of economic theory's core premises. And yet, to people outside of the economics profession, one of the basic problems with economics is that its assumptions seem to be highly unrealistic. There are two potential difficulties here. One is that, even if the postulates of economic theory could be taken as truths, the acknowledged fact that

---

[236] Robbins (1935)
[237] Schumpeter (1949), p. 33. See also Machlup (1978), p. 468.
[238] Koopmans (1957), p. 142.
[239] Machlup (1978), p. 183.
[240] Hausman (1992), p. 207.

they only address a partial set of the factors at work in economic phenomena means that they can only describe the workings of a counterfactual world. Second, it is a matter of some dispute that the postulates of economic theory can be regarded as irrevocably true; and would appear to most thoughtful people as self-evidently false, or at best as first approximations. As the leading logical empiricist, Ernest Nagel, put it in his classic 1961 book *The Structure of Science*:

"The discrepancy between the assumed ideal conditions for which economic laws have been stated and the actual circumstances of the economic market are so great, and the problem of supplying the supplementary assumptions needed for bridging this gap is so difficult, that the merits of the strategy [of economics as an *a priori* science] in this domain continue to be disputed."[241]

Amongst the most controversial of the 'ideal conditions' referred to above by Nagel is the idealised notion of Economic Man. The limitations of Economic Man are not difficult to highlight. For example, the assumption that humans are capable of correctly ordering all their possible consumption preferences requires a perfect knowledge of current conditions and a sufficient understanding of the future as to make it accurately predictable. Some may find this unreasonable or conceptually infeasible. As Frank Knight put it 'ignorance, error and prejudice in innumerable forms affect real choices'[242]. To Knight, the immeasurable nature of uncertainty made it inevitable that individuals will make forecasting errors that are incompatible with their assumed ability to rank all preferences and choices. He also argued that the desired ends of individuals were subject to unpredictable change, and were often not individual but inherently 'social' in a way that economic theory failed to address ('competitive sport plays a role at least as great as the endeavour to secure gratifications mechanically dependent on quantitative consumption'[243]); and that the process of trying to determine our wants is actually a major part of human activity ("We strive to 'know ourselves', to find out our real wants, more than to get what we want. This fact sets a first and most sweeping limitation to the conception of economics as a science."[244]).

The concern with the self-evidence of the premises of an economic theory increase as we move from the most basic behavioural postulates to more ambitious assumptions. This issue becomes clear when two alternative theories of the same phenomenon use contradictory postulates as the basis for describing and explaining the behaviour of the phenomenon. Take the behaviour of the savings rate as an example. According to classical economic theory, the amount that people are willing to save is an increasing function of the rate of interest. (And the intersection of this function with the demand from firms for investment capital determines the rate of interest). This may seem plausible and intuitive, after all a higher interest rate implies a greater reward can be obtained from saving. But the implications of this behaviour may seem perverse to the actuary, who is trained that a higher interest rate means that less needs to be reserved today to fund a given sum for tomorrow.

By contrast, in Keynesian theory of interest, the amount people save is not a direct function of the rate of interest. Rather, Keynes specified the behaviour of saving only indirectly by what is left over after consumption has taken its share of income. His postulated law for the behaviour of consumption was 'that men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income'[245]. This, Keynes argued, was a 'fundamental psychological law, upon which we are entitled to depend with

[241] Nagel (1979), p. 509.
[242] Knight (1956), p. 167.
[243] Knight (1956), p. 269.
[244] Knight in Tugwell (1924), p. 229.
[245] Keynes (1936), p. 85.

great confidence both *a priori* from our knowledge of human nature and from the detailed facts of experience'[246].

The point here is not to say one theory is better than the other. It is merely to point out that there exist basic postulates of economics that appear to be mutually contradictory and yet are each regarded as self-evident truths by some economists. It may, however, be said of the natural sciences that many mutually incompatible theories happily co-exist. After all, scientific knowledge is a 'patchwork of laws'[247] rather than a single unified theory of everything. However, economic theory has not reached the degree of consensus typical of the mature physical sciences. Mutually incompatible economic theories occupy the same piece of the patchwork of laws, competing for legitimacy. And, as we have noted above, empirical corroboration of either theory may not be forthcoming.

Where does all this leave us? Doubt about the self-evident truth of economic's basic postulates critically undermines the rationale of the *a priori* approach: if the postulates can only be considered as approximations to reality, then in the absence of empirical testing of the theories' predictions, how do we know that the approximations to reality that the postulates represent are good enough (*ex ante*) to predict given real-world economic phenomena? All scientific laws have their limitations, a point at which they are no longer accurate or valid. In the absence of empirical testing of a theory's predictions, how do we know where this limit exists for economic theories (other than *ex post*)?

The crux of the problem is that, if we cannot take the postulates of economic theory as self-evidently true or self-evidently sufficient to adequately capture the relevant drivers of real economic phenomena, economic theories that are intended as positive statements then must be tested like other positive scientific theories: by empirically testing their deduced consequences.

Deductive Theoretical Systems in Economics – Empirical Testing of the Predictions of Economic Theories: Part 1 (Hutchison)

The previous section established that some support for a form of *a priori* method has been present in economic thought throughout the development of economics since the mid-nineteenth century. But serious concerns about its use as a core part of economic methodology started to emerge in the 1930s. The most notable contribution of this period was made by Terence Hutchison in his 1938 book, *The Significance and Basic Postulates of Economics*.

Hutchison was highly critical of the then-prevailing methodology of economic theory, regarding its use of vague *ceteris paribus* clauses and reliance on *a priori* reasoning as tautological, unfalsifiable and, hence, unscientific. His book was the first to explicitly advocate Popperian falsificationism for the methodology of economics (Popper's *Logic* was first published, in German, in 1934). Hutchison urged economists to focus on developing theories that could produce empirically testable, falsifiable predictions. He argued this must involve recognising 'inextricably interconnected social phenomena' by working more closely with other areas of social science such as sociology and political science. Popper's influence in this regard is clear, but Hutchison is less clear about how economics can overcome the well-recognised inevitable practical difficulties involved with the empirical testing of its predictions. Whilst his exaltation for economics to take heed of important contemporary developments in philosophy of science was well-received by much of the economics profession who wished to be seen as in tune with sophisticated modern methodological thought, the falsifiability criterion implied that much of mainstream economic theory was essentially unscientific.

---

[246] Keynes (1936), p. 85.
[247] The phrase was coined by Nancy Cartwright, the contemporary philosopher of science.

Moreover, and more controversially, Hutchison's empiricism actually went beyond Popper's. Hutchison urged not only that an economic theory's predictions are subject to rigorous empirical testing. He also argued that a theory should be assessed by the direct empirical testing of its postulates and assumptions. This focus on the realism of the assumptions of the theory may sound similar to a non-dogmatic form of *a priorism*. But it differs in a crucial respect: *a priorism* works from the basis that the assumptions are self-evident and therefore do not require empirical testing; to require the assumptions to be empirically tested as part of the testing of a theory is a methodologically distinct argument. It really ran counter to most contemporaneous philosophy of science - even ardent realist philosophers would agree that the scientific theory should be assessed by the empirical performance of its predictions, not its assumptions. Furthermore, the basic postulates and assumptions of an economic theory may be no easier to test than the theory's predictions. The postulates could only be easily tested if non-observable, theoretical terms had no role to play in economic theory. Hutchison's methodology therefore reduces economic theory to the identification of descriptive empirical relationships between directly-observable phenomena. Causal laws and explanation had been effectively banished.

Hutchison's position was doubtless inspired by Bridgman's operationalism, which was published a decade earlier (see Chapter 1.1 for a brief discussion of operationalism). Operationalism was quite a fashionable doctrine of philosophy of science in the 1930s. It lost favour, however, as philosophers recognised the important role that non-observable terms played in successful scientific theory. Thus, by the 1950s, the logical empiricism of philosophers such as Braithwaite, Nagel and Machlup rejected operationalism and embraced the model of a hierarchy of hypotheses where higher-level hypotheses containing theoretical non-observable terms played an important role in scientific theories (see Chapter 2.1).

Nonetheless, Hutchison's book was highly influential amongst economists and its influence endured for decades (though it seems that what economists may say about methodology and what they practice can be two quite different things). Fritz Machlup, the logical empiricist with a particular focus on economics, wrote an article in 1956 that dismissed Hutchison's requirement for the empirical testing of a theory's basic postulates and assumptions as a form of 'ultra-empiricism', writing that 'the test of the pudding lies in the eating and not in its ingredients'[248]. That is, the testing of scientific theories should be concerned with the empirical performance of its predictions, not its assumptions. But, as Machlup himself clearly recognised, unfortunately it was often much harder to 'eat' the theories of economics than those of the natural sciences.

Whilst Hutchison's work may not have effected a revolution in the methodology of economic theory, it did awaken economists to the prospect that their methods may seem scientifically lacking, especially in light of the major developments in philosophy of science that were taking place during this period. And changes in the methodology of economics were afoot which were broadly aligned to at least some aspects of Hutchison's empiricism. Quantitative methods developed quickly during the 1930s-1950s period. As we saw in Chapter 1, this was a period of major development in statistical inference. Much effort was put into improved collection of macroeconomic statistical data, and econometrics, that most empirically focused of economic disciplines, started to emerge as a major sub-discipline of economics around this time. Econometrics is discussed more specifically below in Chapter 4.4.

[248] Machlup (1956). See also Machlup (1978), Chapter 23.

Deductive Theoretical Systems in Economics – Empirical Testing of the Predictions of Economic Theories: Part 2 (Friedman)

Milton Friedman wrote the twentieth century's most influential essay on the methodology of economics. It was published in 1953, in the midst of the era of logical empiricism. Friedman was one of the most significant economists of the twentieth century and the winner of the Nobel Prize in economics in 1976. Whilst being enormously influential, the vast secondary literature that Friedman's methodology essay spawned included a very significant amount of criticism from both philosophers and other economists.

It is unclear whether Friedman was aware of contemporaneous developments in the philosophy of science (his paper did not cite any philosophers), but his essay was broadly aligned to the major tenets of the school of logical empiricism. Friedman argued for a unity of method across natural and social sciences. He emphasised that the purpose of all science was to make empirical predictions and that the only way of testing the worth of a theory was by the empirical testing of its predictions:

"Its [positive economics'] task is to provide a system of generalisations that can be used to make correct predictions about the consequences of any change in circumstances. Its performance is to be judged by the precision, scope and conformity with experience of the predictions it yields."[249]

And, more emphatically: "the only relevant test of the validity of a hypothesis is comparison of its predictions with experience."[250]

Friedman was more optimistic than most about the prospects of testing the predictions of economic theory. He argued that the evidence provided by experience is 'abundant and frequently as conclusive as that from contrived experiment' though he conceded the empirical evidence in economics is 'more difficult to interpret…frequently complex and always indirect and incomplete'. Thus, Friedman accepted that the empirical testing of the predictions of theories is more difficult in the social sciences than in natural sciences, but nonetheless held that it remains the only way of testing a theory, and that such testing can often be sufficiently achieved in the case of economic theories.

So far, this is quite orthodox and consistent with the contemporaneous logical empiricist's view of how to test hypothetico-deductive theoretical systems. In emphasising that the only purpose of science was empirical prediction, he also expressed the view that the truth and realism of a theory's assumptions were irrelevant. This is also not philosophically radical: it is the instrumentalist perspective that was discussed in Chapter 2.4. It is not inconsistent with a positive scientific method or with the prevailing logical empiricist orthodoxy. To recap, realists argue that theories are at least approximately true representations of reality; instrumentalists argue no one knows if scientific theories are true, approximately true or completely untrue, but that it does not really matter, as what matters is a theory's ability to reliably and accurately predict the behaviour of some particular phenomena.

Friedman's unabashedly instrumentalist view, however, provoked considerable controversy amongst economists and philosophers. There are perhaps a few likely reasons for this reaction. Friedman, in making full use of his considerable rhetorical skills, may have overstated his case. Indeed, he argued that realism in assumptions was not merely irrelevant, but that *less realism may tend to be preferable to more*. That is, that the best, most powerful theories are simple ideas that can explain much of real-life complexity with little. But, again, simplicity is one of the most accepted desiderata

[249] Friedman (1953)
[250] Friedman (1953)

of the scientific method, and it seems possible to view Freidman's perspective here with more sympathy than it was granted. But it did leave Friedman open to the accusation that his methodological argument was actually principally motivated by a desire to defend neoclassical economic theory (and some of its outlandishly simple postulates).

Another factor in the level of criticism that Friedman's instrumentalism attracted is likely to be that instrumentalism as a philosophical doctrine generally moved increasingly out of favour relative to realism amongst philosophers of science over the course of the twentieth century. Finally, as we have seen above, economic methodology up until the publication of Friedman's paper had tended to defend its methods using a highly realist perspective. That is, economic theories and their behavioural postulates had typically been defended as self-evidently realistic descriptions of the true workings of economic phenomena (the limiting case being a form of *a priorism*). Friedman's instrumentalism was about as far from the *a priorism* that characterised the century of economic methodology that preceded his paper as it was possible to reach.

Inevitably, a theory or model with highly unrealistic assumptions can only have predictive success within some limited domain. This is itself is not a fatal issue – a limited domain of applicability must ultimately exist for any scientific theory. And Friedman acknowledged this and argued that the theory's predictive success should only be judged for 'the class of phenomena the hypothesis is designed to explain'. But how do we know the boundary conditions of validity for the theory? Isn't it trivial or tautological to say a successful theory makes empirically supported predictions within those boundary conditions? Isn't it the point of scientific theories to make novel and unexpected predictions well beyond the phenomena being directly considered by the theory? These sort of arguments explain why realism has tended to win out over the instrumentalist perspective over the last several decades.

But whether realist or instrumentalist, the most fundamental difficulty with a methodology for economics as a positive science is that the empirical testing of economic theories' predictions is so practically difficult. Friedman's optimistic perspective on the empirical testability of the predictions of economic theory is a fundamentally different one to that of economists such as Mill, Schumpeter, Robbins and Knight, who, as discussed above, argued that the inherent uncertainty and variability in the set of causes at work at any moment in time generally rendered the testing of economic theory by empirical economic observation of effects impossible. Contemporary writing on economic methodology continues to recognise the limitations for empirical testing of the predictions of economic theory, and there is a broad consensus that the prospects for such testing are significantly limited. As a leading contemporary economic methodologist has noted, 'strict adherence to falsificationist norms would *virtually destroy all existing economic theory*'[251]. We will return to what this may mean for the status of economic theory in Chapter 4.2.

Experimental economics

The above discussions have been driven by the major challenges in performing empirical tests of economic theories' predictions. This, in turn, arises from the inability to perform controlled, repeatable experiments in the social sciences. In recent decades, a major field of economics has emerged called 'experimental economics'[252]. It has become a well-established branch of economics, and one of the leaders in the field, Vernon Smith, received the Nobel Prize in economics in 2002. Can it deliver new capabilities for the empirical testing of economic theories' predictions? Alas, the answer is only in some quite narrow domains – it may be possible to create laboratory conditions for

[251] Hands (1993) in Hausman (2008)

[252] Smith (1994) provides a useful overview.

testing efficient auction design or a specific prediction of game theory, but it is not possible to test Keynes' liquidity preference theory in such a way.

Interestingly, experimental economics seems to have been mainly focused on testing the postulates of neoclassical economic theory rather than the theory's deduced predictions. This is, in essence, Hutchison's 'ultra-empiricism'. It at least provides some answer to the charge of dogmatism that may be levelled at *a priorism*. That is, economists cannot be accused of unscientific dogmatism if they are open to using experimental techniques to test their basic postulates rather than maintain that they are self-evident truths. (However, that can only be said if economists are willing to change their assumptions when they are demonstrated to be false, which, as we will see below, is not so clear.)

Experimental economics has, however, generated some significant results – perhaps most notably, several studies suggest types of human behaviour that are incompatible with the postulate that humans are rational utility-maximising individuals[253]. The responses of the economics community when faced with these contradictions to its self-evident postulates has generally been quite defensive. They have raised a couple of key arguments: first, economists have only ever argued that the postulates are first approximations, and the occasional contradiction is therefore not consequential; second, there is no better theory available than that produced by the standard behavioural postulates of neoclassical economics. But, as has been argued above, when faced with the reality that the postulates are only approximations to truth, positive science requires the theory's empirical predictions to be tested in order to understand whether these approximations are indeed consequential or not. It is this aspect of economics as a positive science that has proved so difficult to implement with any success, and so far the field of experimental economics has been unable to alter this state of affairs.

Contemporary appeals to realism

In grappling with the 'unfalsifiability' of economic theory, contemporary economic theorists (such as Robert Sugden[254]) and philosophers of economics (such as Uskali Maki[255] and Tarja Knuuttila[256]) have appealed to the realism or 'credibility' of a theory's postulates as grounds for making an 'inductive leap' from the assumed world of the theory to the real world.

A starting point for this line of defence is that things are not so simple even for the natural sciences. We saw in Chapter 2 how the Duhem-Quine thesis showed that there was good reason to doubt that any scientific theory could ever be empirically falsified (at least, by a single observation). And Maki argues that even the theories of the natural sciences must necessarily be unrealistic[257]. In the theories of physics, the regularities that are identified are specified in idealised conditions that require unrealistic or even unrealisable assumptions about the conditions under which the regularities hold (Newton's mechanics ignores the effects of any frictional forces, classic electromagnetism assumes a perfect vacuum, etc.). So, both natural and social sciences build theories that develop quantitative relationships in idealised conditions that do not fully and accurately represent the real world or that isolate a sub-set of particular causal factors of interest.

---

[253] Lichtenstein and Slovic (1971) is the canonical example, and its methodological implications are discussed at length in Hausman (1992), Chapter 13.
[254] Sugden (2000)
[255] See, for example, Maki (1994) for an overview of his perspective on realism and economic theory.
[256] Knuuttila (2009)
[257] Maki (1994)

A characteristic difference between the two, however, is that in the natural sciences these idealised conditions can often be sufficiently approximated in experimental conditions such that empirical testing of the theory's predictions is permitted. Or, the theory can be sufficiently 'de-idealised' using already-accepted scientific theory so that it can correspond to the available experimental conditions, again allowing the theory's predictions to be tested (and potentially falsified). In economics, on the other hand, the things that are being assumed away are not well-understood extraneous causes that are subject to their own well-understood laws such as is often the case in the natural sciences. For example, the assumption that economic agents behave rationally is not merely the isolation of a cause, it is deliberately assuming away another potential cause (irrational or erroneous behaviour) whose effect on the theory is not understood or measurable[258].

As a result, the 'inductive leap' is surely bigger in economic theory than it is for the established theories of physics. Nonetheless, it is a leap that some argue can be justified. As Sugden puts it: "models…describe credible counterfactual worlds. This credibility gives us some warrant for making inductive inferences from model to real world".[259]

But this sounds distinctly similar to *a priorism*: the notion of an 'inductive leap' based on 'credible' assumptions appears to be saying little more than that the theory's postulates appear quite realistic, and this provides a justification for using the theory even though its empirical predictions are untestable and the theory is unfalsifiable. A leap of faith is required to take the relationships identified in the special simplified world of the theory and apply them to the more general real world.

Some other leading contemporary philosophers of social sciences such as the Cambridge-based Tony Lawson have pursued programs of philosophical realism that have concluded that the methodology of economic theory has failed to deliver a successful positive science and is incapable of doing so:

"Modern economics mostly fails to illuminate the world in which we live and, indeed, is in a state of disarray….we ought to do something about it, and specifically seek to replace, or at least supplement, dominant strategies with others that are rather more capable of being explanatorily successful and useful."[260]

Perhaps the only way out of this rather unsatisfactory state of affairs is to note that the reasonableness of Sugden's logic must depend on *what the theory's inductive inferences are used for*. For example, are the 'inductive inferences' intended to be predictions, accurate to 4 significant figures, of, say, the effect of an immediate change in the Bank of England's inflation target from 2.5% to 3.0% on the level of youth unemployment in North Wales in 2025?[261] Or is the inductive inference intended to be of a more qualitative form, where the theory provides some otherwise non-obvious *insight* into the nature of the expected effects of a given cause?

The case for the latter appears much stronger than the case for the former. For example, the Black-Scholes-Merton[262] model tells us that in a world where perfect dynamic replication of an option is possible, the risk premium of the underlying asset is irrelevant to the arbitrage-free option price.

---

[258] See Knight (1935).
[259] Sugden (2000)
[260] Lawson (1999)
[261] Most of this discussion has used examples from microeconomics. It is worth noting that the size of the 'inductive leap', this challenges of mapping a theory to empirical reality, is greatest when the scope of the theory is widest. Macroeconomics, the study of the functioning of the whole economy, is therefore naturally the field in which the consequences of these challenges can be observed most acutely.
[262] Black and Scholes (1973); Merton (1973).

This result is not (to me, anyway) an obvious consequence of the premises of the theory. We know that perfect dynamic replication of an option is not possible in the real world. But the striking and non-obvious insight provided by the theory may nonetheless make it reasonable to judge that changes in the underlying asset's risk premia will not have a first-order effect on option prices (*ceteris paribus*). This insight *has not been produced by the positive scientific method* as described in Chapter 2. Indeed, the model manifestly fails as a predictor of empirical behaviour (option prices do not have constant implied volatilities that remain constant over time, as is implied by the model). *It is an interpretative form of knowledge that has been arrived at through 'conceptual exploration' of a counterfactual 'ideal-type' world*.

In a nutshell, and in the language of Chapter 3, should economic theory be considered as a positive or interpretative form of theory? The answer may be both. There may be circumstances and types of problem where economic theory can function positively in delivering accurate, reliable quantitative empirical predictions in the way Friedman claimed. But the above discussion suggests that, more often than not, the real value in economic theory is the form of abstract qualitative insight that it is capable of delivering. We next further explore this promising idea of economics as an interpretivist, as opposed to positive, discipline.

4.2 Beyond positivism….an interpretivist perspective on economics

Chapter 4.1 highlighted the profound methodological challenges inherent in the development of economics as a successful positive science. We concluded that the issues around the performance of economics as a positive science may be resolved by recognising it is not a positive science, but an interpretative discipline. That is, the function of economic theory is not to deliver accurate and reliable empirical predictions, but to provide conceptual insight that can aid our understanding of the workings of empirical reality.

The most fundamental issue that recurred in our discussion of economics as a positive science was the difficulty of obtaining testable empirical predictions from economic theory. This difficulty may be considered to arise (at least in part) from the partial nature of economics: it isolates a sub-set of the motivations of human behaviour and can therefore only explain and predict the empirical behaviour driven by that sub-set of factors. Economic theory also features other forms of idealisation, simplification and abstraction which creates a significant gap in correspondence between the postulated world of the theory and the reality of a given human society.

Why have economists chosen to do this? Primarily, because it makes their theory tractable and manageable. The real economic world is one of overwhelming complexity and continuous change. This raises a natural question: is this a complexity that can be progressively mastered as economic theory and techniques improve? Put another way, is it reasonable to expect that economics as a positive science is simply at a relatively early stage of development, and it can be expected to incrementally progress in the way that positive science is anticipated to in a Popperian model of the growth of scientific knowledge? Or are there fundamental epistemic barriers to this development that arise from the nature of the phenomena that economics considers that makes such efforts futile?

It would seem reasonable to assume that the answer must be that the relationships between *some* phenomena *are* constant and uniform across a very wide range of conditions and circumstances, and hence are capable of accurate and reliable quantitative description and prediction in the ways associated with a successful positive science. But the limited empirical predictive success of much of economic theory suggests that the relationships between many economic phenomena and the human behaviour that drives them are so complex and contrary to the methodological premises of positive science that the methodology of positive science cannot be reasonably expected to master them.

Knightian risk and uncertainty

Frank Knight's 1921 book[263] famously distinguished between measurable risk and immeasurable uncertainty. This distinction offers a fundamental reason for why a positive scientific method may fail in areas of economics (and indeed other social sciences).

Knight distinguished between three types of situations in which probabilities could arise when predicting the future behaviour of variable phenomena in the world of business and economics. They can be summarised as follows:

- Type 1: The phenomenon's probability distribution and its parameters *are known* from prior knowledge or the use of some general (non-empirical) principles. No reference to empirical observation is therefore necessary to determine the probability distribution. The phenomenon is defined in such a way that all instances of it are homogenous. Put another way, its characteristics are known to not change over time or across instances.

- Type 2: The phenomenon's probability distribution is unknown. A sample of observations from the population of instances of the phenomenon is available. The population is known to be homogenous and its characteristics are known to not change over time. The sample is known to be independent and identically distributed (or exchangeable in subjectivist probability terminology).

- Type 3: The phenomenon's probability distribution is unknown. It is not known if the population of instances of the phenomenon is homogeneous and / or we do not know if its characteristics are stable through time. Sample observations may or may not be available. However, it they are available, it is *not* known if the sample data are independent and identically distributed (exchangeable).

Knight classified Types 1 and 2 as *measurable risk*; and Type 3 as *immeasurable uncertainty*. The classification as three distinct types suggests clear boundaries between them, but Knight emphasised that, except at the idealised extremes, there was really a continuum of differences in degree rather than differences in kind.

Various aspects of these probabilistic settings were discussed in Chapter 1. Let us take the canonical example of an urn of black and red balls to explore Knight's definitions. The Type 1 case corresponds to the setting where the number of balls of each colour in the urn are known. In the Type 2 case, the number of balls of each colour in the urn is not known, but a sample of random selections (with replacement) from the urn have been made and the results of that sample are known. In such a scenario, both objectivist and subjectivist theories of probability will ultimately arrive at similar conclusions from the given empirical data in making forms of valid statistical estimates of the ball colour proportions with measurable statistical errors. In Type 3, there may be absolutely no information about the number and colour of balls in the urn, other than that each ball can only take one of two colours, red or black. The discussions of Chapter 1 highlighted that this setting could cause much philosophical angst. Subjectivists who subscribe to the Principle of Indifference would say the probability of the next ball randomly taken from the urn being red is one half, the same as

[263] Knight (1921)

the probability of it being black. Objectivists argue that it is absurd to claim a logically valid numerical probability exists in these circumstances.

This simple urn example highlights how Knight's types can differ by degree. For example, does a sample of one ball denote a shift from Type 3 to Type 2? And, as we saw in Chapter 1, an objectivist would *define* the probability (in a Type 1 setting) according to what the sample probability tended to as the (Type 2) sample size tended to infinity. So, this suggests we can steadily move across a spectrum from Type 3 to Type 2 to Type 1 as the (independent, identically distributed or exchangeable) sample size increases from 0 to infinity.

In the Type 3 setting, some other form of indirect information about the composition of the urn may be available beyond the results of a sample. For example, perhaps there are two empty tubs of red paint and one empty tub of black paint sitting on the floor beside the urn. Knight argued that, in a general business and economic context, this is actually the sort of information that most commonly arises (that is, business decisions are typically taken in a unique set of circumstances and thus cannot be considered as belonging to a homogenous set, but nonetheless are supported by informed analysis). Knight described the process by which qualitative information is used in the estimation of the probabilities of uncertain (Type 3) events in the following terms: "The ultimate logic, or psychology, of these deliberations is obscure, a part of the scientifically unfathomable mystery of life and mind."[264] Thus, to Knight, Type 3 probability estimation is a judgement-based estimation and is *subject to immeasurable error*.

What does this have to do with the methodology of economics? Knight argued that standard neoclassical economic theory could accommodate risk (Types 1 and 2), but not uncertainty (Type 3)[265]. Unsurprisingly, this is not a view that has been universally accepted by positively-minded economists. Milton Friedman, for example, argued that Knightian uncertainty was handled perfectly well by subjective probability, and that subjective probability was fully compatible with neoclassical economic theory[266]. But this seems in danger of missing the point: the issue at stake is not the internal logical coherence of an economic theory that assumes individuals behave consistently with their subjective probability evaluations; rather, the difficulty is with the inductive challenge that may arise when applying a theory whose deductions require phenomena (such as the dynamic behaviour of asset prices) to follow Type 2 processes (probability distributions of prices are known risks) when in reality they follow a Type 3 process (the asset price's stochastic process cannot be known).

Knight further argued that some people were better than others at estimating Type 3 probabilities in their chosen fields (through the use of intuition, expertise and judgement) and that this skill was the essential source of entrepreneurial profit in competitive markets. Thus, his thesis argued that the basic postulates of orthodox economic theory were missing an important element, and that this element should be explicitly incorporated into the theory. But how can immeasurable uncertainty be captured by a methodology of positive science that is based on regularities and uniformities? In a nutshell, it cannot. *The distinction between risk and uncertainty marks a limiting boundary for positive science.*

[264] Knight (1921), p. 227.

[265] We may also note, for actuarial interest, that the models of financial economics generally work with Type 1 probabilities. In the standard models of portfolio theory, option pricing theory, and interest rate term structures, the means, variances, correlations, stochastic processes and probability distributions are all taken to be exogenous inputs that have no estimation uncertainty.

[266] Friedman (1976), p. 82-84.

Economics and Interpretivism

Whilst Knight argued that the empirical presence of immeasurable uncertainty rendered much of economic theory inadequate as the basis of a positive science, he nonetheless explicitly argued in support of the value and usefulness of 'pure' deductive economic theory[267]. But its use would require a degree of judgement that was beyond that required of a positive science. In his Presidential Address to the American Economic Association in 1950, he explicitly rejected economics as a positive science and advocated an interpretivist alternative methodology:

"*The formal principles of economic theory can never carry anyone very far toward the prediction or technical control of the corresponding behaviour*….the intelligent application of these principles is a *first step*, and chiefly significant negatively rather than positively, for showing what is 'wrong' rather than what is 'right' in an existing situation and in any proposed line of action. Concrete and positive answers in the field of economic science or policy depend in the first place on judgements of value and procedure, based on a broad, general education in the cultural sense, and on insight into human nature and social values, rather than on the findings of any possible positive science. From this point of view the need is for an interpretative study."[268] [italics added]

The term 'interpretative study' is itself open to interpretation. What might an interpretivist methodology for economics look like? We saw in Chapter 3.3 that Weber, the early champion of interpretivism as a methodology for social science, focused on the analysis of unique and significant 'concrete' historical events rather than theoretical deductive systems. Does this mean that the interpretivist view of the huge body of work represented by deductive economic theory must be that it is meaningless and useless? Knight's answer to this was an unambiguous and emphatic 'no'. Rather, to Knight, economic theory and economic models may be powerful when used as what Weber called an 'ideal type'. This was certainly (and unsurprisingly) also the view of Max Weber:

"Pure economic theory…utilizes ideal type concepts exclusively. Economic theory makes certain assumptions which scarcely ever correspond completely with reality but which approximates it in various degrees and asks: how would men act under these assumed conditions if their actions were entirely rational?"[269]

Fritz Machlup, writing in 1964, offered a very similar perspective to Knight on the use of economic theory within an interpretative methodology:

"Economic theory is based on counterfactual assumptions, contains only theoretical constructs and no operational concepts, and *yields results which, we hope, point to elements of truth present in complex situations*."[270] [italics added]

In more contemporary times, Daniel Hausman made a distinction between economic theory and economic models. Whilst theory was intended to make direct statements about the behaviour of the empirical world, economic models were "definitions of kinds of systems, and they make no assertions. It is a category mistake to ask whether they are true or to attempt to test them. Their point lies in conceptual exploration and in providing the conceptual means for making claims that can be tested and can be said to be true or false."[271]

---

[267] See, for example, Knight (1921), Chapter 1, especially p. 6.
[268] Knight (1956), p. 177.
[269] Weber (1949), p. 43-44.
[270] Machlup (1978), p. 483.
[271] Hausman (1992), p. 78.

This perspective on the purpose of economic models is clearly consistent with a Weberian interpretive methodology that views such models as ideal types (a point which Hausman directly acknowledges[272]). And, as was noted in Chapter 4.1, the contemporary philosophical output on inductive leaps and credible counterfactual worlds appears most naturally resolved by putting it firmly in the context of an interpretivist rather than positivist methodology. Indeed, *the entire issue of* a priorism *and the difficulty in testing a theory's empirical predictions largely disappears when the deductive theory is viewed as an interpretative model or ideal type rather than as a positive scientific hypothesis whose predictions should be directly empirically tested*. Recalling the discussion of Chapter 3.3, considering a theory or model as an interpretative 'ideal type' (as opposed to a positive scientific theory) means that its function is *not* to provide direct prediction or explanation of empirical phenomena, and the usefulness of the theory or model should not be measured by the extent to which it does so. This has major implications for what constitutes a successful model.

The interpretative model provides a purposeful abstraction from the real world, one which highlights (deductively) how particular phenomena will behave under specified conditions. The interpretivist acknowledges that the ideal type's postulates may be conceptually far removed from the real world, and that the theory's predictions may differ materially from behaviour in the real world. Nonetheless, it can provide powerful insight into the form of interrelationship and interaction of certain types of phenomena in the real world. It does not provide quantitative empirical predictions. And, ultimately, *the extraction of useful real world insight from the interpretative model requires analytical skill, expertise and judgement (subjective interpretation) of a different form to that employed in the positive scientific method.* This use of analytical skill, expertise and judgement is the interpretivist solution to the inductive leap from counterfactual theory to empirical reality. The solution is less than perfect – what exactly does skill, expertise and judgement constitute? – but the leap is arguably less daunting when the target is the generation of insight for experts instead of accurate and reliable empirical prediction.

Treating an economic theory or model as an interpretative study using ideal types rather than as a positive scientific hypothesis resolves much of the methodological difficulties discussed in this chapter. We noted at the end of Chapter 4.1 how the value of the Black-Scholes-Merton model could be much better appreciated when viewed as an interpretative model rather than a positive scientific theory. This conclusion may similarly apply to much of the quantitative theory of financial economics (especially in the field of asset pricing more generally, for example, yield curve models and capital asset pricing models). The theories of financial economics do not directly specify behavioural postulates for individuals, and so the interpretative notion of 'inner understanding' or 'sympathetic empathy' that was an important element of Weber's interpretivism is not present. Nonetheless, the models of financial economics are highly abstract idealisations, consequently often have dubious performance in empirical prediction, and yet can be powerfully insightful and useful as interpretative ideal types.

To take another famous example in the field which should be familiar to actuaries, Vasicek's original paper[273] on the arbitrage-free modelling of yield curves is enormously enlightening as a study of how a stochastic model specification for the short-term interest rate, together with exogenous risk premia assumptions, can determine the arbitrage-free, stochastic dynamics of an entire yield curve through time. The model specification is highly abstract and counterfactual. It does not provide a remotely realistic description of the empirical behaviour of short-term interest rates or yield curves

---

[272] Hausman (1992), p. 80.
[273] Vasicek (1977)

(a fact Vasicek readily acknowledged in his paper). But the paper provided completely new insight into how the stochastic process for the short rate acts as a causal factor on the behaviour of long-term bond prices in the absence of arbitrage. It demonstrated, for example, how a higher degree of mean-reversion in the stochastic process for the short rate could translate into lower volatility in the arbitrage-free prices of long-term bonds. It is an abstract economy theory that is highly effective as an interpretative study in the sense described by Knight, Machlup and Weber above.

Moving beyond financial economics, a similar argument can be applied to many areas in the core of neoclassical economic theory. General equilibrium theories, which are highly abstract attempts at modelling the simultaneous determination of all prices in an economy are a good (and arguably notorious) example[274]. In all of these cases, the power of the theory lies not in its ability to predict empirical phenomena such as the pricing of assets, but in its ability to provide deep insights into the right and, especially, the wrong (see the Knight quotation above) ways to think about economic processes.

Much of this chapter's discussion essentially resolves into an analysis of the *purpose* and *use* of an economic theory or model. The conclusion of this discussion has been that economic reality is sufficiently complex that, except in very particular circumstances, economic theory cannot be relied on to provide reliable accurate quantitative empirical predictions. That does not mean that economic theories or models have no useful role to play. But the role is essentially *heuristic*. *The theory or model can provide an expert with new insight into a complex problem, enhancing their understanding and aiding their assessments and decisions. But this is only possible when the theory or model's output is interpreted with a deep understanding of its capabilities, assumptions and limitations*.

Alas, the hunger for evidence-based and 'scientific', objective answers has meant that economic models are increasingly directly used as the means of quantitatively determining the 'right answer', rather than as an interpretative aid for experts. This effect can be observed throughout the social sciences. As we shall discuss in Part II of this work, it has also been the root of a notable trend in the evolution of some key areas of actuarial practice in recent decades.

4.3 Econometrics and macroeconomics

The essence of econometrics is well-captured by the definition provided by the US economist and econometrician Arnold Zellner in 1984: econometrics is "the field of study in which economic theory, statistical methods, and economic data are combined in the investigation of economic problems"[275]. It might also be added that the economic problems that econometrics tackles tend to be in the field of macroeconomics rather than microeconomics.

There is relatively little academic literature that is focused on the philosophical issues surrounding econometrics specifically. But econometrics can be regarded as the 'the principal empirical method of economics'[276] and so the philosophical arguments that pertain to positivist economics apply equally, and indeed with particular vigour, to econometrics. Chapter 4.1 and 4.2's discussions of the philosophy and methodology of economics are therefore directly relevant to econometrics. We shall try to avoid unnecessary repetition.

Nonetheless, there have been some interesting methodological and philosophically minded debates about the practice of econometrics and its prospects of success. These debates have mainly (but not

[274] Hausman (1992), Chapter 6.7.
[275] Zellner (1984), p. 306.
[276] Hoover (2001), p. 144.

exclusively) taken place amongst leading economists and econometricians rather than between philosophers. A brief review of some of these debates will highlight the point that is obvious from our review of the methodology of positive economics: the difficulties involved in making empirical predictions in a world that is infinitely complex, non-stationary, and (therefore) characterised by significant immeasurable uncertainty is the fundamental recurring philosophical theme in econometric methodology.

A History of Underwhelming Performance

Econometrics started to emerge as a recognised specialist discipline in the 1920s and 1930s in the slipstream of the major developments in statistical inference pioneered by Ronald Fisher[277]. Work on the collection of national-level economic statistics began in earnest in countries such as the UK and US in the nineteenth century, and this was at least partly driven by a desire to empirically test economic theory. The collection of such data underwent serious institutional development in the western world in the 1920s – both in terms of the breadth of statistical quantities that were collated and the quality of the data gathering. This was at least partly inspired by the empirically focused American institutionalist school, which was instrumental in establishing the National Bureau of Economic Research in the US in 1920. Its work was crucial in providing the data that fuelled econometric models (even though the data were often highly aggregated with non-homogenous underlying constituents). Meanwhile, the Great Depression created a demand for new economic policy solutions, further stimulating interest in a more empirical approach to economics.

Early econometricians aspired to marry economic theory and empirical data analysis in a way that would elevate economics to the methodological standards expected of a positive science. When the Econometric Society was founded in the early 1930s, one of its stated aims was:

"to promote studies that aim at unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking similar to that which has come to dominate in the natural sciences."[278]

The empirical testing of the predictions of economic theory was therefore one of the key objectives of early econometrics, and it remains so today. However, economic theory rarely, if ever, implies a fully specified form of econometric model, especially in macroeconomic theory. The econometrician, in attempting to test theoretical predictions, is therefore required to perform a wider role than merely applying obvious statistical tests to economic data. As Julian Reiss, the contemporary philosopher of economics put it:

"Econometrics does not only apply statistical methods to pre-existing economic theories. Rather, it transforms theoretical claims into testable empirical propositions that can be expressed in a precise mathematical form and statistically estimates their parameters."[279]

But this is no straightforward task. Many feasible systems of empirical equations may be consistent with a given economic theory and yet result in conflicting conclusions. Consequently, it is difficult (and therefore unusual) for econometrics to deliver empirical evidence that can 'falsify' a given economic theory. Mutually contradictory theories can therefore compete alongside each other, seemingly indefinitely, moving in and out of relative fashion as fads and philosophical tastes dictate. For example, monetarism (money supply causes price changes) and real business cycle theory (prices cause changes in money supply) offer contradictory causal models of fundamental macroeconomic

---

[277] For a full discussion of the early history of econometrics, see Morgan (1992).
[278] Frisch (1933)
[279] Reiss (2013), p. 162.

phenomena, and econometric analysis of the theories' relative empirical performance has been ambiguous at best[280].

The empirical testing of economic theory, however, has not been the only objective of econometrics. Other objectives include making reliable macroeconomic forecasts and predicting the impact of government policy changes on the behaviour of economic phenomena. In the task of economic forecasting, econometric models have a long history of poor performance. Since the late 1940s, leading economists such as Milton Friedman have queried whether the complexity of econometric models really adds any value to their predictive performance[281]. And analysis of the predictive performance of these models has often failed to determine that they perform any better than the simplest back-of-the-envelope modelling techniques[282].

Econometric models have also endured similar criticism of their ability to predict the effects of changes in economic policies. Academic analysis of the behaviour of competing econometric models has found that "they disagree so strongly about the effects of important monetary and fiscal policies that they cannot be considered reliable guides to such policy effects…"[283]

The critical and consistent point in all of the above criticism is that the methodological challenges of econometrics arise from its intrinsically empirical nature. It is the tip of the spear of economics as a positive science. To the sceptic of positive economics, the disappointing results we have briefly outlined above would come as no surprise.

Non-Stationarity

The basic philosophical difficulty with econometrics is that its methods of forecasting and predicting are largely inductive. They fundamentally rely on an assumption of a uniformity of nature that is difficult to reconcile with the complex, continually changing economic world we live in. Continuous and unpredictable changes in the characteristics of the underlying macroeconomic processes hobble the applicability of sampling theory and statistical inference, as such changes leave us with *a single observation of the path of history that cannot be reliably de-composed into many independent, identically distributed well-behaved little pieces*.

The history of econometrics is replete with examples of problems created by long-term non-stationarity. To take an early examples, in 1921, when W.H. Beveridge, the then director of London School of Economics and to-be political architect of the British welfare state, analysed four centuries of wheat price data, he noted that the behaviour of the time series fundamentally changed in the nineteenth century. He attributed this to 'the disturbing influence' of industrialisation and the credit cycle and concluded the effect 'cannot be eliminated by any simple model'[284].

Unsurprisingly, this basic concern with the inherent difficulty of predictive modelling of non-stationary economic systems has echoed through the decades. Examples of expressions of this concern from leading economists of the twentieth century are not particularly hard to find. In his 1970 Presidential Address to the American Economic Association, Wassily Leontif noted:

"In contrast to most physical sciences, we study a system that is not only exceedingly complex but is also in *a state of constant flux*. I have in mind not the obvious changes in the variables, such as

---

[280] See Hoover (2001), Ch. 10.
[281] Zellner (1984), p. 101.
[282] See, for example, Kamarck (1983), Chapter 7.
[283] Christ (1975); Zellner (1984), p. 111, for fuller discussion.
[284] Beveridge (1921), p. 432; see Morgan (1992), p. 34 for fuller discussion.

outputs, prices or levels of employment, that our equations are supposed to explain, but the basic structural relationships described by the form and the parameters of these equations."[285]

Sir John Hicks, the leading British economist and recipient of the 1972 Nobel Prize in Economics, wrote in his 1979 treatment of causation in economics:

"The more characteristic economic problems are problems of change, of growth and retrocession, and of fluctuation. The extent to which these can be reduced into scientific terms is rather limited; *for at every stage in an economic process new things are happening, things which have not happened before – at the very most they are rather like what has happened before*."[286] [italics added]

The problem of non-stationarity is just as pervasive in econometrics today. For example, a 2019 paper[287] on the empirical estimation of the Phillips curve (the relationship between the rate of unemployment and the rate of inflation[288]) noted how statistical estimates of the slope of the curve had materially flattened over the last several decades. The authors suggested that a significant part of the explanation was likely to be found in changes in how individuals form long-term inflation expectations. These in turn had been caused by structural and previously unanticipated changes in the monetary policy of the US Federal Reserve in the early 1980s. But research highlighting changes in the empirical slope of the Phillips curve has been produced since 1970, if not before[289]. Some of the various explanations offered include changes in the age-sex structure of the labour market, changes in unemployment benefit and changes in trade union power[290].

The presence of significant non-stationarity in economic systems suggests that using empirical historical data to develop probabilistic predictions of the economic future may be a fool's errand. It might, however, be argued that non-stationarity merely means the model can only be expected to work well 'locally' in time: that short projection horizons (over which sufficient stationarity in conditions can be anticipated) can be more reliably predicted than long horizons; and that recent historical data is more relevant than distant historical data. But if it is accepted that a given system is exposed to the unquantifiable effects of material unanticipated exogenous shocks, then it is hard to see how any projection horizon can be the subject of reliable probabilistic prediction.

Moreover, if only recent data is deemed relevant for calibration purposes then this will tend to result in inadequate data sample sizes for use in the calibration and testing of econometric models (bearing in mind these will often be complex non-linear models featuring hundreds of parameters). The data may only include observation of a fairly limited set of economic circumstances. It may not allow the model to robustly extrapolate beyond those circumstances to the ones that may matter for the purposes of risk assessment or policy decision-making.

The Lucas Critique

The Lucas critique was noted in Chapter 3.2's discussion of positivism in the social sciences. Whilst its logic has wide applicability across all social science, it was developed in the specific context of macroeconomics, and so deserves a further mention in this discussion. In the context of econometrics, the Lucas critique can be viewed as highlighting a special kind of non-stationarity: if a particular set of government policies prevailed during the historical period to which the econometric

[285] Leontief (1971), p.3.
[286] Hicks (1979), p. xi.
[287] Hooper et al (2019)
[288] Phillips (1958)
[289] See, for example, Schultz (1971).
[290] Roll (1992), p. 534.

model was calibrated, there is no reason to necessarily expect the model to be valid under a different set of policies.

The Lucas critique implies that an econometric model cannot be reliably used to assess the effects of new hypothetical policies. This logically undermines the use of econometric models in the evaluation of economic policy choices. We saw an illustration of this effect in the Philips curve example given in the non-stationarity example. A change in central bank policy on inflation targeting meant that the prior relationship between inflation and unemployment was materially altered.

The degree to which the Lucas critique applies depends on the extent to which the model captures the underlying 'permanent' features of behaviour, rather than merely describing recently-observed empirical relationships that are a function of more fundamental behaviours. Lucas' recommended solution was therefore to advocate using 'bottom-up' models of individual decision-making processes that can more directly capture the causal relationships that economic policy is attempting to impact upon. But these fundamental causal factors are often unobservable phenomena such as economic agents' expectations and the adjustments that these agents will make when faced with new government policies[291].

Econometrics and Causation

The difficult relationship between theory and measurement that arises in econometrics is exemplified by the history of how econometricians have wrestled with the concept of causation[292]. Earlier discussions have highlighted that causation, like induction, has a tendency to become philosophical quicksand.

The American post-war economist Herbert Simon defined causation purely with reference to the relationships specified by an abstract deductive model, and not in terms of the 'real-world' empirical relationship of the phenomenon (or the empirical predictive performance of the model)[293]. At the other end of the philosophical spectrum, the British economist, Clive Granger, proposed a definition of causation that was entirely based on the statistical relationships in empirical data, without any reference to a law or other form of causal explanation[294]. Simon's definition of causation might be referred to as 'theory without measurement' whilst Granger's is 'measurement without theory'. Both of these approaches therefore miss at least one of the two crucial aspects of causation as discussed in Chapter 2.3 (recall how Hempel and Oppenheim argued that a scientific explanation of cause and effect required a logical deduction from a general or universal law that has empirical content.)

Moreover, different econometric analyses of the 'Granger causality' present between macroeconomic variables such as the money supply and gross national product have arrived at diametrically opposite conclusions[295]. This perhaps reflects the difficulties of working with small sample sizes and the inevitable low power of the statistical testing of the models and calibrations that use them (a property which econometricians may have tended to underappreciate[296]). Low statistical power amplifies the limitations of measurement without theory: if there must be a

---

[291] See Hoover (1994) for a more extensive methodological discussion of this 'identification problem' in econometrics.
[292] See Zellner (1984), Chapter 1.4, for a fuller discussion of the history of econometrics and causation.
[293] Simon (1953)
[294] Granger (1969)
[295] Zellner (1984), p. 61.
[296] Zellner (1984), p. 279.

tendency to be stuck with whatever null hypothesis we start with, one that is derived from economic theory is presumably a better start than one that has been arbitrarily selected.

Tinbergen and Keynes

A discussion of the methodological debates in econometrics would not be complete without a brief reference to the famous historical dialogue between Jan Tinbergen and John Maynard Keynes.

Tinbergen, a Norwegian economist, wrote two papers on business cycle modelling for the League of Nations in 1939[297]. He is regarded as one of the great pioneers of econometrics and empirically focused macroeconomic modelling. His business cycle modelling, the most ambitious and extensive of its time, developed separate macroeconomic stochastic models for the US, UK and Dutch economies. The models specified stochastic behaviour for consumers, labour markets, corporations, banks and financial markets in 50 stochastic modelling equations containing hundreds of parameters, which were calibrated to available historical data. Tinbergen's model was intended to be consistent with macroeconomic theory, but his specification of the workings of so many parts of the economy inevitably involved subjective modelling choices. Many of these were made with a primary consideration towards mathematical and statistical convenience.

Keynes, who we saw in Chapter 1 was capable of deep philosophical thinking on probability, was highly sceptical of Tinbergen's application of statistical techniques to empirical economic data[298]. Whilst this criticism and the debate it sparked is now some eighty years old, the topics are fundamental and timeless and continue to concern econometric methodology today.

The main strands of Keynes' argument were that it was not possible to identify the full model and all its relevant factors or their functional dependencies; that many of the factors were unobservable and hence immeasurable; and that the model structure could not be expected to remain stable over time. As we have seen above, these are all arguments that other leading economists and philosophers have made in their criticisms of econometrics over the last one hundred years.

Tinbergen went on to share the first Nobel prize in economics (with Frisch) in 1969 for his contributions to econometric modelling, so clearly not all economists shared Keynes' scepticism. Keynes' view was not that all empirical analysis in economics was pointless. He conceded that quantitative models could be useful tools for describing and providing insight into what had gone before. His fundamental philosophical criticism was concerned with the conversion of this historical description into an inductive inference about the future. Keynes also argued that labouring under the pretence that this form of inductive inference was reliable was not only futile, but destructive:

"In…natural sciences the object of experiment is to fill in the actual values of the various quantities and factors appearing in an equation…and the work when done is once and for all. In economics that is not the case, and *to convert a model into a quantitative formula is to destroy its usefulness as an instrument of thought*."[299] [italics added]

The argument that quantitative predictive modelling is not only futile but destructive is the most fundamental form of warning against positivism in the social sciences. Perhaps surprisingly, in the second part of this book, we will find this is a view that has been expressed by at least one of the 20th century's most important actuaries.

---

[297] Tinbergen (1939)

[298] See Lawson and Pesaran (1985), Chapter 8 for a fuller discussion of Keynes' views on econometric modelling and his critique of Tinbergen's work.

[299] Keynes (1973), p. 299.

Part Two: On the Methodology of Actuarial Science

5 Contemporary methodology of actuarial science

Part One considered some of the major topics of philosophy of science that may be potentially relevant to the methodology of actuarial science. These topics included the philosophy of probability and statistics; inductive inference from the observed to the unobserved; the scientific method and how it differs in the natural and social sciences; positivist and interpretivist perspectives on method in social science; the philosophy of economic theory and econometrics; and the measurability of the risk and uncertainty associated with the phenomena of the social sciences.

Part Two considers what lessons the philosophical analyses of these subjects may offer for actuaries. The actuarial profession is a learned profession. It is concerned with the professional application of technical knowledge to fulfil the functions that lie within the profession's domain of expertise. This technical knowledge we will, for our purposes and to align with convention, call actuarial science. It will often be an application of other, independently developed, 'pre-existing' science, most often originating in the domains of mathematics, statistics, finance and economics. It will, however, also make use of technical knowledge developed in increasingly diverse fields such as computer science, demography and medicine. It is not the original source of this scientific knowledge, but *its application to questions recognised as within the actuarial domain of expertise*, that we will use to determine what constitutes actuarial science.

This chapter will aim to first *identify some key characteristics* of the technical knowledge that is representative of actuarial science. This is challenging because actuarial science uses such a wide range of techniques and methods from a variety of fields. Nonetheless, there is some commonality to the types of questions that actuaries seek to answer, and this helps to determine some fundamental characteristics for the way the profession makes use of the diverse library of scientific knowledge at its potential disposal. Having identified these general characteristics of actuarial science, we will then reflect on what major methodological considerations they naturally give rise to, based on what has been surveyed in Part One. This analysis will inform our view of which methodological approaches are likely to make for successful actuarial science and which are not. Chapter 6 will then consider the implications of Chapter 5's analysis for how the methodology of actuarial science may further develop. First, Chapter 5.1 briefly highlights some of the significant historical discussions of the methodology of actuarial science that have already gone before us.

5.1 A brief history of thought on the methodology of actuarial science

Historically, research on the philosophical underpinnings of actuarial science has been relatively thin on the ground. Some of the historical highlights of actuarial output on methodology are outlined below.

A historical summary of the methodology of actuarial science may begin in the 1760s with Richard Price. Price played a uniquely interesting role in the development of both inductive inference and actuarial science. It was Price who was bequeathed Thomas Bayes' unpublished musings on probability and statistics upon Bayes' death in 1761, and who saw that they were published by the Royal Society (of which both Bayes and Price were both Fellows) in 1763. If it wasn't for Price, the historical development of Bayesian probability and statistics may have been quite different. It would certainly have been called something different (Laplacian, probably). From 1768 to the 1780s, Price was a consultant to Equitable Life, and helped to establish the financial architecture of the British mutual life office and the engine of its growth, the with-profits policy. It was a structure that was replicated by many other life offices in the decades that followed, and which was successfully sustained for over 200 years.

Price's major actuarial publication was the 1772 book, *Observations on Reversionary Payments*. Price was a technically sophisticated intellectual, but his actuarial ideas and methods were uniformly diffused with practical awareness and commercial vision. He was concerned with how life assurance policies should be fairly priced, how mortality rates could be estimated from population data, how life offices could be run in a financially secure and sustainable way that was equitable across generations and across different classes of policyholder. His actuarial output was pragmatic, empirical and applied. He did not suggest his actuarial work was in search of universal truths or new types of fundamental scientific knowledge. His mortality tables were rigorous analyses of historical data, but he was the first to include the caveat that 'at particular periods, and in particular instances, great deviations will often happen'[300].

Not all historical actuaries retained Price's intellectual humility. Gompertz published an article in 1825 proposing a parametric formula for how mortality rates vary by age[301]. He deduced this formula from a premise about how a human's ability to fend off death would deteriorate with age: 'the average exhaustion of a man's power to avoid death were such that at the end of equal infinitely small intervals of time, he lost equal portions of his remaining power to oppose destruction which he had at the commencement of those intervals.'[302] He found that when his formula's parameters were fitted to some standard mortality tables, the formula could provide a reasonable fit to the mortality table's numbers over a large portion of table's age range. When Gompertz discussed his paper decades later at the International Statistical Congress of 1860, he argued his formula represented a 'law of human nature' that he had deduced from fundamental principles[303]. Gompertz saw himself as a pioneering scientist, finding permanent laws of uniformity on a par with physical laws of nature. He elaborately compared his law of human mortality to the physical mechanical system of a piston in a tube producing air condensation.

Senior actuaries of Gompertz's era such as such as Jellicoe, Makeham and Woolhouse all advocated this philosophically ambitious interpretation of Gompertz's formula. By the end of the nineteenth century, however, such an interpretation was strongly disputed by some within the actuarial profession, perhaps because the observations of mortality experience that had since been accumulated could highlight the less than permanent nature of Gompertz's law and its parameters.

Thomas Young, in his Institute Presidential Address of 1896, provided a particularly robust critique of Gompertz's ambitious claims[304]. Young argued that medical developments and societal changes meant that no stable law of human mortality rates could be present over time. He also argued that, to be regarded as a law of nature, the quantification of the relationships in the formula should not only be stable over time, but should be deduced directly from the theory (and then subject to empirical tests), rather than merely be calibrated to some empirical data. Young's points would seem very reasonable in the eyes of much of the modern philosophy of science discussed in Part One (especially Chapter 2). Nonetheless, it is doubtless the case that in many fields of empirical social science, such as econometrics, empirical study is focused on the parameterisation of a theory rather than on the testing of a theory that has theoretically deduced parameter values.

[300] Price (1772), p. 129.
[301] Gompertz (1825)
[302] Gompertz (1825), p. 518.
[303] Gompertz (1871)
[304] Young (1897). See also Young (1880).

Little of note seems to have been written by actuaries explicitly on methodology or philosophical considerations during the first half of the twentieth century. This is perhaps surprising given this was a period of much activity and significant progress in the philosophy of science and probability (albeit this was largely focused at the time on the physical rather than social sciences). However, by the 1950s, some articles were starting to appear in actuarial journals that reflected on some of these developments, particularly in the philosophy of probability and the topic of the relative merits of the objectivist and subjectivist approaches to probability (the subject matter of Chapter 1 above).

The famous mathematician and Bletchley Park veteran, I.J. Good, published an article in the *Journal of the Institute of Actuaries* in 1956 in which he advocated the actuarial adoption of Bayesian methods[305]. A couple of notable papers on the actuarial application of Bayesian techniques were published in the American actuarial profession's journals in the mid-1960s[306]. In 1975, a polemical article in favour of the objective perspective on probability and statistics was published in the *Journal* by the Australian actuary J.M. Ryder[307]. We noted in Chapter 1 how a 1981 paper by Ryder published in the *British Journal for the Philosophy of Science* made him the somewhat accidental architect of the intersubjective approach to subjective probability (he was vehemently anti-subjectivist in philosophical outlook, and his paper on intersubjective probability was actually intended as a refutation of subjective probability, rather than as its potential salvation). Ryder's 1975 paper reflected his strong objectivist persuasion. With the possible exception of the American papers on Bayesian applications, it is not obvious that any of these papers excited great interested amongst the actuarial profession. The *Journal* papers of Good and Ryder were somewhat abstract and theoretical, and their potential implications for actuarial methods would likely have been unclear to the typical practicing actuary.

The 1980s and 1990s saw some quite notable material published on actuarial methodology. In particular, Frank Redington, in the closing stages of his career in the 1980s, made some strong statements about actuarial methods and techniques that suggested a deep philosophical scepticism towards positivist methods in the social sciences in general, and in actuarial science in particular[308]. And in the late 1990s, the actuary John Pemberton presented a substantial and controversial paper on the methodology of actuarial science that was published in the *British Actuarial Journal*[309]. The Staple Inn discussion of this paper was opened by Nancy Cartwright, the highly renowned philosopher of science. A couple of years earlier, Paul Huber completed his PhD thesis on the philosophical foundations of actuarial economic models[310]. We will refer again to this material of the 1980s and 1990s as we proceed through Part Two. Interestingly, most of it presented a scepticism about the methodology of positivism in social science that was at odds with much of social science practice of the time, and that was increasingly at odds with how actuarial methodology had been developing since the 1970s.

We will now attempt to identify the key general characteristics of actuarial science as a scientific discipline. This will allow us to identify which different themes of philosophy of science may have particular implications for the methodology of actuarial science. Given this characterisation of

---

[305] Good (1956)
[306] Jones (1965) and Jones & Kimeldorf (1967)
[307] Ryder (1976)
[308] Redington (1986), Section 8.
[309] Pemberton (1999). Chapters 2 and 3 of this paper cover similar ground (more concisely) as Chapters 2 and 4 above. Pemberton was an actuary at Prudential plc in the 1990s and then became a scholar of philosophy of science at London School of Economics.
[310] Huber (1996)

actuarial science, we will then be in a position to determine the most significant areas of methodological consideration that arise and their potential implications.

5.2 General characteristics of actuarial science as a scientific discipline

Let us briefly re-cap Part One's discussion of the scientific method and its application in the natural and social sciences. From there we will then attempt to characterise the nature of actuarial science and find its natural place in this methodological framework.

A re-cap of some key ideas from Part One

In Chapter 2, the hypothetico-deductive method was presented as a key model of the logic of the scientific method. This method describes empirical relationships between phenomena of interest via scientific laws that are deduced from an axiomatic description of the relevant phenomena. Sophisticated and mature sciences may have a hierarchy of laws, with the most general and theoretical high-level hypotheses being used to deduce many wide-ranging lower-level hypotheses that are each capable of direct empirical testing. Empirical support for the higher-level hypotheses is derived from the empirical success of lower-level hypotheses. Empirical support for the lower-level hypotheses is indirectly derived from the empirical success of other lower-level hypotheses that belong to the same hierarchy of theories. The more diverse and numerous the empirically successful lower-level hypotheses, the greater the support for the higher-level theory and each of the various lower-level hypotheses. These hierarchical systems of theories can be considered as one interpretation of Kuhn's idea of a paradigm, but this terminology was not used by the logical empiricists who fully developed the hypothetico-deductive method in the 1950s.

Depending on philosophical taste, these scientific laws can be interpreted as a description of (or at least a good approximations to) reality and truth (*realism*), and hence provide a form of *causal explanation* for the joint behaviour of particular phenomena; or the laws can merely be regarded as a form of efficient and useful quantitative *description* that renders apparently complex interactions into simpler relationships (*instrumentalism*).

The theory of the hypothetico-deductive method matured in the 1950s in the hands of a school of philosophers of science known as the logical empiricists. We saw in Chapter 2.5 that a post-positivist protest against some of the tenets of the logical empiricists' hypothetico-deductive method emerged in the 1960s and 1970s. This work highlighted that the hypothetico-deductive could only be regarded as an abstract and simplified model of the scientific method. But post-positivism failed to produce a superseding model of how scientific knowledge is produced. The hypothetico-deductive model remains a good basic foundation from which to understand and describe how scientific knowledge is incrementally developed in the natural sciences.

The greatest methodological complications of the hypothetico-deductive method arguably arise when the scientific theory includes probabilistic behaviour for some phenomena. This has increasingly become the norm in 20th and 21st century science: quantum theory, the most successful theory of the physical sciences of the last century or so, is irreducibly probabilistic; and much of the social sciences, with its descriptions of the aggregate behaviour of humanity, are also often couched in probabilistic terms. The hypothetico-deductive method can accommodate probabilistic statements, but it muddies the logic and creates a new source of fallibility in empirical testing. The 'replication crisis' of modern scientific research is undoubtedly exacerbated by the ubiquitous presence of probability statements in scientific theories and empirical generalisations. 'False discovery rates' are hard to avoid when statistically testing the empirical evidence for theories that do not have a high prior probability (in Bayesian terminology) of being correct. Such low prior probabilities can arise either from testing bold conjectures in a Popperian sense (see Chapter 2.1); or from the vast and indiscriminate searches of thousands of potential relationships that can be undertaken by modern machine learning algorithms (discussed further in Chapter 6.3).

In Chapter 3, we found that much of social science has thus far failed to reach the sophisticated state of scientific knowledge described in the above discussion of the hypothetico-deductive method. The social sciences have not often been able to develop a theoretical framework that can be described as a large hierarchical structure of logically related hypotheses of varying degrees of generality and direct testability. In many social science disciplines, empirical work has focused mainly on identifying and quantifying low-level empirical relationships rather than deducing testable theories that can predict and explain causal relationships.

Chapter 3 noted an important philosophical divide with major methodological implications for the social sciences: there was a school of thought, that can be traced back to Comte's late 19th century publications, that argues that the empiricism and positivism of the hypothetico-deductive method can be applied to the social sciences in a fundamentally similar way to that of the natural sciences; and there is another school of thought, less homogeneous than the positivist school, but with recognisable roots that go back at least as far as Weber's work of the early 20th century and arguably much further, that is sceptical that such a method can achieve success in many fields of social science. This scepticism reflects the argument that the phenomena of social science exist in an infinitely complex and intrinsically non-stationary environment that makes accurate quantitative prediction of their future behaviour practically very difficult or even conceptually impossible. We saw in Chapters 3 and 4 that a quite broad array of 20th century thinkers adopted a form of this position (Weber, Keynes, Knight, Hayek, Popper, etc.). This argument leads to a rejection of a positivist method in social science. That is, theories of the social science should not be expected to generate reliably accurate empirical predictive success. This does not imply that social science is worthless or that it has no legitimate role for theory. Rather, it leads to the idea that theories of social science are useful as abstract interpretative studies that can provide deep and valuable insight in the hands of experts who understand the theories and their inevitable epistemic gaps with empirical reality.

Chapter 4 focused on economics as an important and methodologically interesting specific field of social science. Economics is a field of social science that actuarial science has attempted to make increasing use of over the last fifty years. Economics is methodologically interesting because it is one of the very few fields of social science that has developed axiomatic theoretical frameworks that appear capable of use as part of a hypothetico-deductive method. However, the discussion of Chapter 4 found that the economic theory's application of the hypothetico-deductive method had distinctive features when compared to its application in the natural sciences. Economic theory only ever attempts to provide a *highly partial* description of reality. As a result, it is usually impossible to empirically test the theory – its quantitative predictions do not apply to our empirical reality and it is generally not possible to artificially create the experimental conditions where the predictions do apply. This, together with a general subscription to the scepticism of social science positivism that was discussed in Chapter 3, led us to the conclusion that economics should not be regarded as a positive science in the sense that it is capable of reliably accurate quantitative predictions. The chapter did, however, emphatically endorse economics as a powerful interpretative study, and identified examples where actuarial science could gain deep and practically useful insight when economic theory is used in this way.

The anatomy of actuarial models: cashflow projections and discount functions

Let us now consider actuarial science and where it may fit into the above methodological schema. As argued above, actuarial science encompasses a wide array of methods, techniques and knowledge, and the boundaries of actuarial science should be determined by the type of questions that are being considered rather than by the specific content of the strategy used to answer those questions. These questions almost invariably involve some form of financial assessment that is based on the projected behaviour of some phenomena (such as mortality rates, inflation rates or corporate defaults) that a relevant future event (a term assurance pay-out, the size of a pension payment, the size and timing of a bond asset casflow) is contingent upon. (Note that, throughout this work, when we write of a model's projected behaviour for some empirical phenomena, we do not attempt to make a distinction between the meanings of verbs such as project, predict and forecast.)

The most fundamental activity of actuarial science is to quantitatively assess the present value of a series of predicted future cashflows. Actuarial models' output can usually be ultimately expressed as the calculation of such a present value. The cashflow projection and the discount function that is applied to them may be deterministic or stochastic. Actuaries may also build models that project how those present values will behave in the future. Again, this projection may be deterministic or stochastic. Most actuarial calculations can ultimately be cast in these terms.

The calculation of present values, however, is of course not the privileged domain of the actuarial profession. Virtually all of financial theory and practice – from arbitrage-free derivative pricing theory to a clothing firms' appraisal of whether to invest in the design of a new range of swimwear – can be viewed as a present value calculation. Are there typical properties of the present value calculation that are found in actuarial models? To tackle this question, it may be useful to think about the present value processes that characterise virtually all actuarial work in a couple of distinct, albeit related, dimensions. First, there is the nature of the future cashflows and, in particular, *what sort of contingency determines the size and timing of the cashflow*; and, secondly, there is the purpose of the present value calculation, which determines *the choice of discount function*.

Typically, the cashflows of an actuarial calculation will have very well-defined contingencies. They may depend on the behaviour of financial and economic variables – inflation rates or future asset values; and they may depend on the health status, death or survival of a policyholder, or perhaps some other policyholder status, such as being made redundant by an employer; or the cashflows may depend on some general form of risk and its consequences (losses due to property damages caused by a hurricane or a motor accident). These well-defined forms of contingency are usually of a form that makes the future cashflows amenable to some form of attempted prediction via quantitative empirical analysis of relevant data. However, almost invariably, the phenomena that actuaries attempt to predictively model will also be similarly modelled by experts outside the actuarial profession. Demographers also model and predict human mortality. Economists also model and predict inflation rates. Meteorologists model and predict hurricanes. Rating agencies model and predict corporate defaults.

What about the discount function? Are there forms of discount function that uniquely exist in the privileged professional domains of actuaries? Again, not really. There are some forms of present value that actuaries calculate and use that have a distinctive actuarial flavour, and some that are similar or identical to the present values that are assessed by other financial professionals. For example, the present value may be intended to assess the fair value of a financial assets (say a loan or a mortgage) and in this case the valuation methodology and the purpose of the valuation will likely be very familiar to other financial professionals such as accountants, economists and investment analysts. Nonetheless, the combination of the type of contingency together with the purpose of the present valuing may create a distinctive actuarial role. For example, the fair valuation of a mortality-contingent asset such as a lifetime mortgage would generally involve a significant role for actuaries even though most of the required technical expertise could arguably be provided by a combination of accountants (fair valuation expertise), demographers (mortality modelling expertise) and economists (expertise on the valuation of contingent economic cashflows).

There are also many examples of other present valuing processes in actuarial science that have a more distinctive actuarial flavour: the assessment of a prudent amount of capital to hold to support the long-term risks on an insurance balance sheet; or the amount of assets that a Defined Benefit pension fund should hold today such that the fund will, on average and under a given investment strategy and contribution policy, meet its future pension liability cashflows, for example. These processes may involve a logic and method that is less recognisable to non-actuarial finance professionals. But here again we may note non-actuarial parallels. For example, some forms of financial institution (such as banks) have assessed prudential capital requirements for centuries without any actuarial role being required.

It is useful for methodological purposes to consider the two fundamental parts of the actuarial modelling process, i.e. the projection, prediction or forecasting of future cashflows; and the conversion of those cashflows into some form of present value metric; as distinct tasks. They are often related: for example, the purpose of the present valuing may determine whether the cashflow forecast should be made on a best estimate basis or with some margin for prudence; in some special cases (such as option valuation), the approach taken to projecting cashflows (the assumed size of risk premia) may uniquely determine the discount function to be applied. But, in general, the function of assessing those predicted cashflows and the function of converting them into some present value metric are separate activities that require different skills and will make use of knowledge and theory from different disciplines.

Most discussions of the methodology of actuarial science and the characteristics of actuarial models have focused heavily on the first of these functions: on the projection of future cashflows (and / or the projection of asset or liability values). This projection function is, of course, a fundamentally important and defining actuarial activity. But the second of these functions, the conversion of a set of (deterministic or stochastic) projected cashflows into some form of present value measure, is equally important to the output of actuarial analysis. This part of the actuarial process is fundamentally driven by an unambiguous definition of the question that the actuarial analysis is intended to answer. This can involve some subtlety and, above all else, clarity of thought. It has been the greater source of complexity and confusion throughout the history of actuarial thought, and it's a topic that will recur in our discussions of the methodology of actuarial science, particularly in Chapter 6.

The key distinctive characteristics of actuarial science according to Pemberton and Cartwright

In his notable (and sometimes controversial) actuarial methodology paper[311] of the late 1990s, John Pemberton summarised his characterisation of actuarial models with three key words: "*empirical, applied and approximate*"[312]. The discussions of the scientific method in Chapter 2 highlighted that all of science can really be characterised by these terms. Pemberton's point, of course, was that these properties were present in the methodology of actuarial science to a greater degree than was the case for most other scientific disciplines. Nancy Cartwright, notable as one of the most influential philosophers of science of the last quarter of the twentieth century, opened the Staple Inn discussion of Pemberton's paper. She offered an alternative, but complementary, and perhaps sharper three-word summary of the distinctive characteristics of actuarial science: "*local, bottom-up*

---

[311] Pemberton (1999)

[312] Pemberton (1999), p. 177.

and skill-based"[313]. We now attempt to unpack what philosophers of science mean when they describe actuarial science in such terms.

Actuarial science applies knowledge, concepts and ideas from both the natural and social sciences. It is, in the main, concerned with phenomena related to individual human behaviour (which is estimated by attempting to group people into homogeneous groups); and with phenomena influenced by the aggregate behaviour of human society (future economic conditions and financial asset prices, for example). Contingencies such as weather and long-term human longevity trends might be offered as examples of phenomena from the natural sciences that are of interest to actuarial science, but even these cases are clearly influenced by the effects of human society, especially over the long-term. So, at its core, actuarial science would seem to be a social science.

Actuarial science is not, however, primarily concerned with the development of new scientific knowledge from the study of the social behaviour of humans. Actuarial science is the body of technical knowledge that is used in the professional activities of actuaries. Theoretical knowledge – for example, scientific models of causation that can *explain* empirical relationships - is therefore only relevant to the extent it is useful in furthering professional actuarial activities. It is in this sense that actuarial science can be considered an *applied* science. And, in light of the above paragraph, we may go a step further and refer to it as an *applied social science*.

Chapter 3 noted that the hypothetico-deductive scientific method does not generally play a major role in most of social science (with the possible exception of economics, which, as noted in Chapter 4, tends to apply it with its own unique methodological features). The quantitative empirical relationships that social scientists analyse tend not to be developed by means of an axiomatic theoretical system. As an applied social science, it is therefore unsurprising that we should find that actuarial science also does not typically proceed by the use of such systems (Gompertz notwithstanding!). Rather, in actuarial fields such as mortality modelling or the modelling of general insurance risks, the modelling tends to be strongly empirical. That is, the relationships are derived directly from examination of empirical data, rather than being deduced from an abstract theoretical framework and then tested against empirical evidence. *Actuarial science deals with empirical generalisations rather than laws of nature*. The general absence of theoretical explanatory frameworks for the behaviour of the phenomena of interest and the emphasis on immediate empirical relationships gives actuarial science its *empirical* character.

The description of actuarial science as *bottom-up* is closely related to its empirical and applied character described above. Actuarial model-building is bottom-up in the sense that the relationships are built up from low-level inductive generalisations, rather than being developed top-down from the deductions of a higher-level theoretical framework. Actuaries are not searching for universal truths, but merely for efficient and predictively powerful descriptions of the relationships between phenomena of interest (these bottom-up relationships may, nonetheless, be complex and non-linear).

The characterisation of actuarial science as empirical, applied and bottom-up does not imply that actuarial science must reject any use of a causal or explanatory understanding of the phenomena it studies when building actuarial models (though Pemberton came close to suggesting it does). Actuaries may sometimes use model structures that make use of theoretical insights that have usually been generated in other scientific disciplines. Uninterpreted historical data may often be highly inconclusive, and causal explanatory structures, where they exist in reliable form, can be of

[313] Cartwright in Pemberton (1999), p. 178.

much use. ‘Cause of death’ longevity model structures will make use of scientific output from the medical and related professions. The description of the future evolutions of interest rates may make use of arbitrage-free term structure models from financial economic theory. Forms of weather contingent events may rely on climate change forecasts which may be developed using causal theoretical structures from the natural sciences. But it is not the general objective of actuarial science to develop such causal or explanatory theories. In an actuarial setting, they are only useful as a means to another end (but may be highly useful nonetheless).

Our limited understanding of the complex nature of the underlying processes that drive the behaviour of phenomena of actuarial interest, together with limited relevant empirical data, mean that the relationships identified in these models cannot generally be considered as exact. The empirical generalisations that form the basis of actuarial cashflow projections are therefore almost always explicitly recognised as *approximate*. Given the tentative and fallible nature of this modelling knowledge, it is natural that the predictive performance of the model is closely monitored over time, with the actuary ready to make modelling adjustments that reflect the information gleaned from newly acquired experience.

Actuarial work will often be interested in the outcomes for very specific risks. A pensions actuary may be predicting the mortality experience of the pension fund of a Welsh coal mining firm; a motor insurance pricing actuary is ultimately attempting to project the behaviour of the individual driver who may buy the insurance; and like-wise for the life actuary attempting to price individual annuities or life assurance. The pensions actuary may choose to adjust the latest national mortality projections to reflect anticipated differences in the likely longevity of the specific pension fund relative to the pool of lives considered in the wider study. The general insurance or life assurance pricing actuary may use a dozen or more risk factors to attempt to accurately predict the insurance policy’s cashflows (although cost and ethics may restrict the number of risk factors that are used). When a life actuary sets the lapse rate assumption for their company’s assurance product, he or she may choose to use a different assumption from that implied from their recent policyholder experience, because of the anticipated impact of a new change in the surrender value basis. And so on. These adjustments make the models *local*. It is implicit in Cartwright’s discussion that she anticipates these local adjustments can deliver some predictive success.

Often, these differences in assumptions may not be derived entirely from empirical data because the data does not exist in the granularity required. Instead, these adjustments to the actuarial model will be made with the use of judgement and modelling skill. This use of judgement does not fit naturally with the objectivity of positive science. The use of judgement in the building and application of models is what Cartwright is referring to when she says actuarial science is *skill-based*. Pemberton also regarded this as a fundamentally important element of the practice of actuarial science: “The skill of the modeller forms an integral part of the actuarial method”[314]. The choice and structure of the model, the selection of empirical data for use in calibration, and the ‘local’ adjustments that are made to that calibration for use in the specific intended purpose are all important elements of the model-building process that require skill and judgement.

The above discussion has highlighted the main ideas behind philosophers’ use of the descriptive terms empirical, applied, approximate, local, bottom-up and skill-based to characterise the methodology of actuarial science. These descriptions and the ideas behind them (other than Pemberton’s apparent antipathy towards the utility of any theory) are largely uncontroversial - there probably isn’t a great deal in the above characterisation of aspects of actuarial modelling that many

[314] Pemberton (1999), p. 159.

actuaries would find reason to strongly object to. Nonetheless, the uncontentious nature of this characterisation of actuarial science does not imply that it is absent of any significant methodological issues. In particular, as an empirical and applied social science, we cannot avoid a fundamental question: can social or socio-economic phenomena be predicted in a reliably accurate way? The answer to this question dictates the form of knowledge that can be delivered by actuarial models. Can the actuarial model provide a reliable description (probabilistic or deterministic) of the future behaviour of the given phenomena of interest? Or does the deeply complex nature of the phenomena of actuarial interest place significant epistemic limits on the predictive knowledge that it is possible to obtain? In which case, can actuarial models be better viewed as interpretative tools that can provide useful insight? We will explore these questions further in Chapter 5.2 and beyond.

Beyond Cashflow Projections

Before that, however, we will offer one further observation of Pemberton and Cartwright's characterisation of actuarial science and its methodology: it is, in an important sense, *incomplete*. Their characterisation can provide a fine description of the methodological context for the projection of cashflows in actuarial models. It does not, however, address the second key element of the actuarial modelling process – the choice of discount function that is used in the conversion of the cashflows into some form of present value metric.

This part of the actuarial modelling process is not a form of scientific inference that can fit naturally within the field of study of the philosophy of science. It is therefore difficult to make an analysis of the methodological considerations that accompany it, and that is perhaps why it has not been the subject of much philosophical inquiry. The choice of discount function is often not a problem of how to answer an inferential question, but rather of deciding which question is relevant and useful to attempt to answer. Philosophy can help indirectly by helping to inform on the extent to which a given question *can* be reliably answered. But that, in itself, is not sufficient to determine if the answer is useful. This is perhaps ultimately a matter of professional judgment, based on an understanding of what the actuarial client is seeking to achieve.

The choice of discount function is nonetheless fundamental to actuarial output – to its meaning, its usefulness, the questions that it answers and the purpose that it serves. And, as noted above, the crux of the matters concerning most of the historical (and perhaps contemporary) controversies in actuarial science are to be found in this part of the actuarial modelling process.

Chapter 6's detailed discussions of some specific contemporary actuarial issues will reinforce this point. For now, the reader may note how a diverse range of contemporary difficult actuarial topics such as the level of advance funding of Defined Benefit pensions; the definition of insurance risk-based capital requirements; and the fair valuation of illiquid assets, the source of the complexity arises primarily in the choice of discount function that is applied to the projected cashflows rather than in the cashflow projection itself. That is, in all three cases, two actuaries could wholly agree on the projected distribution of future cashflows, and yet come to quite different conclusions as to their actuarial implications for funding, capital or valuation.

5.3 Some key methodological considerations that arise

Several methodological questions naturally follow from the description of the characteristics of actuarial science developed in Chapter 5.2, such as:

- The 'local' nature of actuarial models was emphasised by Cartwright as a fundamental feature of actuarial modelling methodology, and one that may deliver predictive success, at least to an approximate degree. Is this assertion universally reasonable? Are there areas of actuarial modelling where 'localism' cannot ensure (approximate) predictive reliability? Or where the actuarial modelling is necessarily not local? In short, is a belief in the local nature of actuarial modelling as a route to delivering predictive reliability a case of hope over experience?

- Whilst actuarial models may be local with reference to the characteristics of a particular risk or contingency, they are often not local in time. That is, actuarial models may be used to project over very long time horizons. The results of the model may depend on assumptions about the paths of phenomena over very wide spans of time. In the domain of applied social science, with the complexity of non-stationarity that it often entails (See Chapters 3 and 4), can predictive success be reasonably expected over these time horizons? And how does the process of monitoring and updating modelling assumptions work when the assumption refers to what will happen in 20 or 40 years?
- Cartwright described actuarial modelling as skill-based. Can modelling skill be tested or evaluated (*ex-ante* or *ex-post*)?
- Do the axiomatic deductive theories of economics have a role to play in the empirical and approximate world of actuarial science?

These questions summarise the subjects covered by the remainder of Chapter 5.

5.4 Localism and the prospects of predictive success in actuarial modelling

Cartwright's description of actuarial science suggests actuarial models may be able to achieve predictive success (albeit in an 'approximate' way) for socio-economic phenomena in ways that have often eluded other fields of social science. Her characterisation suggests that this success can be obtained in the actuarial domain via two advantages potentially open to actuarial modelling: by limiting the ambition of actuarial models to only strive for 'local' performance; and through the deployment of actuaries' modelling skill and expert professional judgment in the model-building process.

At the same time, Cartwright acknowledged that the philosophy of science has made little progress in understanding or evaluating modelling skill. This raises an obvious question: is such a sanguine view of assured actuarial modelling performance therefore merely wishful thinking? It is certainly tempting to think that many of the sceptics of positive social science that we met in Chapters 3 and 4 – Max Weber, Frank Knight, John Maynard Keynes, Karl Popper and Friedrich Hayek, for example - would believe so. Interestingly, Frank Redington, widely viewed as one of the most important actuaries of the twentieth century, was also firmly of this sceptical view. We now examine more closely what Redington had to say in this regard.

Redington's scepticism about the predictive success of actuarial models

Redington's essay, *Prescience and Nescience*, was written late in his life and published posthumously in 1986[315]. It contains the most explicit exposition of his ideas and views regarding actuarial methodology. Rather like the economist Milton Friedman's venture into the methodology of economics some 35 years earlier, Redington's methodological arguments did not make any explicit reference to any philosopher's work and it is unclear how familiar Redington was with the literature of philosophy of science that had emerged over his lifetime. Unlike Friedman, Redington's perspective on the methodology of social science was strongly anti-positivist. Redington argued, in a similar vein to those such as Knight, Keynes and Hayek, that society and the economy were inherently unstable and subject to continuous change:

[315] Redington (1986). (Redington passed away in 1984.)

"The continued instabilities we observe in our experience are not the regular instabilities of probability theory…the conditions never repeat themselves".[316]

And, he argued, this rendered the behavioural phenomena produced by human society incapable of accurate prediction, and made efforts to attempt to do so futile, as the following quotes illustrate:

"We cannot foresee the future…we cannot foresee forty days let alone forty years."[317]

"The only secure relationship we can have with the long-term future is on the footing that we cannot foresee it. We can of course do a great deal to prepare for it but we cannot foresee it."[318]

"If we cannot foresee the future then we cannot! No proliferation of polysyllables or multivariate analysis will succeed (other than in deceiving ourselves)."[319]

Redington's points here are familiar from the discussion of philosophy of science that was developed in Part One. Inferential techniques, whether simple probability models or more complicated probability models or predictive machine learning algorithms of a sort that Redington had not even envisaged, all rely on an assumption of a uniformity in the nature of the modelled phenomena and the conditions that influence it over time. If that assumption is invalid, the techniques cannot offer reliable quantitative predictive accuracy.

Importantly, however, Redington did not argue that this meant that actuarial modelling was a pointless activity:

"Such an instrument [actuarial model], though it does not help us foresee the future, does the next best thing; it enables us to familiarise ourselves with what the future may have in store. Thus equipped we should be able to home upon our targets with alertness and sensitivity. And that is all we can hope for."[320]

I think Redington is arguing here that actuarial models should be used, in the jargon of Chapters 3 and 4, interpretatively rather than positively. That is, the models should be used to provide useful insights to experts who understand them, and not to attempt to generate the 'true' numerical answer. Redington argued this meant actuaries should use simple models, and he expressed particular scepticism about complex computer-based stochastic models.

I, perhaps unsurprisingly for someone whose career has centred on their application, do not share Redington's particular scepticism about the potential usefulness of stochastic models. Simplicity is always a virtue, whether in the setting of an interpretative ideal type or a positive hypothetico-deductive theory. But there is also always a trade-off between simplicity and the insight that can be delivered for a complex problem. That trade-off is specific to the particular problem at hand. There is nothing special about 'stochastic' in that trade-off. For some problems, a stochastic model may provide interpretative insights in a way that deterministic projections cannot, and the additional complexity the model brings may be a necessary and worthwhile cost. There is no obvious philosophical or actuarial difficulty with that. And just as for any other model (or ideal type), the user must have sufficient expertise and understanding of the model in order for its output to provide them with useful insight.

---

[316] Redington (1986), p. 531.
[317] Redington (1986), p. 518-9.
[318] Redington (1986), p. 520.
[319] Redington (1986), p. 525.
[320] Redington (1986), p. 531.

Let us now move beyond these general methodological discussions and consider some specific evidence. How accurately have major actuarial models predicted key actuarial phenomena? This is clearly a very broad question that could form the basis of a very extensive empirical investigation. For now, we will make do with a couple of examples from some major fields of actuarial modelling.

Actuarial Models and Long-term Predictive Success? Case Study (1): Mortality rate projections

Actuaries have been estimating mortality rates and using these estimates in cashflow projections for hundreds of years. The right to the claim of the first actuarial mortality table is subject to some contention, but candidates would certainly include Halley's 1693 Breslau table and Price's 1772 Northampton table. These mortality tables were based on general population data. As actuarial institutions such as life offices and pension funds developed in scale, actuaries started to use the data generated directly by the experience of their policyholders and /or members rather than the population at large (a simple example of modelling localism in action). By the early 1800s, life offices in countries such as Britain and the United States were pooling their experience data to create very large experience datasets that could support a richer data analysis – for example, allowing the modelling of mortality rates as a function of the duration in-force of the policy as well as the age and sex of the policyholder[321].

Mortality rates have also been used as an illustrative topic in the works of several philosophers of probability. We saw in Chapter 1.1 how the Oxford philosopher A.J. Ayer used mortality rates as an example in his analysis of the reference class problem in frequentist probability. Von Mises, the objectivist, viewed death as a 'social mass phenomena', suggesting it fulfilled his criteria for collectives of uniform and repeatable events[322]. He did allude to the non-stationarity of historical mortality data by accepting 'no figure of this kind, however exact at the moment of its determination can remain valid for ever'. But he also argued this non-stationarity was merely a difference in degree rather than a difference in kind from that found in physical phenomena, noting that even the force of gravity on the earth's surface will change over time[323]. Some philosophers of probability have historically even ventured on to the topic of mortality improvement projection. In 1888, the Cambridge philosopher John Venn, who we met several times in Part One, wrote:

"At the present time the average duration of life in England may be, say, forty years; but a century ago it was decidedly less; several centuries ago it was presumably very much less; whilst if we possessed statistics referring to a still earlier population of the country we should probably find that there has been since that time a still more marked improvement. *What may be the future tendency no man can say for certain*."[324]

The statistical estimation of the mortality rates of a group of people, be it a group of life assurance policyholders or the population at large, is naturally backward-looking. But, as John Venn and Thomas Young pointed out, mortality rates will change over time. Life and pensions actuaries therefore often need to make forecasts of the mortality rates that will be present many decades into the future. Interestingly, until recent decades, this task of quantitatively projecting how today's mortality rates will change over time didn't receive as much attention as the detailed estimation of recent mortality rates. Prior to the final quarter of the 20th century, there was certainly an actuarial understanding that mortality rates were evolving and generally improving over time, but allowance for this future improvement was usually made through some *ad-hoc* adjustments to the-then

---

[321] See Turnbull (2017) Chapter 6, for some more historical discussion.
[322] Von Mises (1928), p. 9-10.
[323] Von Mises (1928), p. 17.
[324] Venn (1888), p. 14.

prevailing mortality rate estimates, rather than through the use of a more formal modelling approach.

The UK actuarial profession's first formulaic approach to mortality improvement projection was published in 1955[325]. Since then, evermore complex approaches to the projection of mortality improvements have been developed. This can be seen partly as an example of the broader tendency towards greater use of more complicated quantitative models in general over this period of history, but also partly as a result of an increasing awareness that, for long-term longevity business such as annuities, the projection of the *change* in mortality rates over the individual's lifetime could be more financially important than an accurate estimate of their mortality rate at the start of the period. Moreover, this importance has been amplified by the changes in economic conditions over recent decades: the financial consequences of forecasting errors in long-term mortality improvement projections are greater in the early 21st century's low long-term interest rate environment than in the higher interest rate environment of the 1960s or 1970s.

Predicting how mortality rates will change over 20, 30 or 50 years is a profoundly difficult challenge. Despite the proliferation of modelling attempts, most actuaries would tend to agree that it is an area of inherent uncertainty and an area where different expert opinions may suggest materially different predictions. But the assumption is unavoidably important to the long-term projection of longevity-dependent cashflows such as those found in Defined Benefit pensions or life assurance annuity portfolios.

Let us now briefly return to Frank Redington and, in particular, his views on mortality rate prediction. Redington wrote extensively on mortality modelling[326]. He believed that there was some physiological or genetic limit to human life, and that this limit may possibly be capable of estimation. But he also believed that environmental factors kept human lifespans far from this limit, and that those factors were 'a clumsy package of very varied articles' such that 'the search for any law other than a rough graduation is in vain'[327]. The anti-positivist leanings he expressed in *Prescience and Nescience* are consistent with this (much earlier) perspective and imply that making reliably accurate long-term mortality rate projections is simply not possible.

Can we examine the actuarial records to determine the historical accuracy of actuaries' long-term mortality predictions? This is perhaps more difficult than it sounds. Firstly, our sample size is very restricted - there are only one or two non-overlapping 30-year historical periods during which actuaries have been publishing explicit quantitative long-term mortality rate projections. So, a formal, rigorous statistical study of the historical performance of long-term actuarial estimates of mortality rates is not easily undertaken.

As an anecdotal starting point, however, we can look back over the last 30 or 40 years and consider what long-term projections have been made and how they have changed. Exhibit 5.1 charts these historical evolutions of long-term mortality rate projections. It shows how the UK actuarial profession's estimate of the mortality rate of a 70-year old in the year 2020 has changed over the last 40 years.

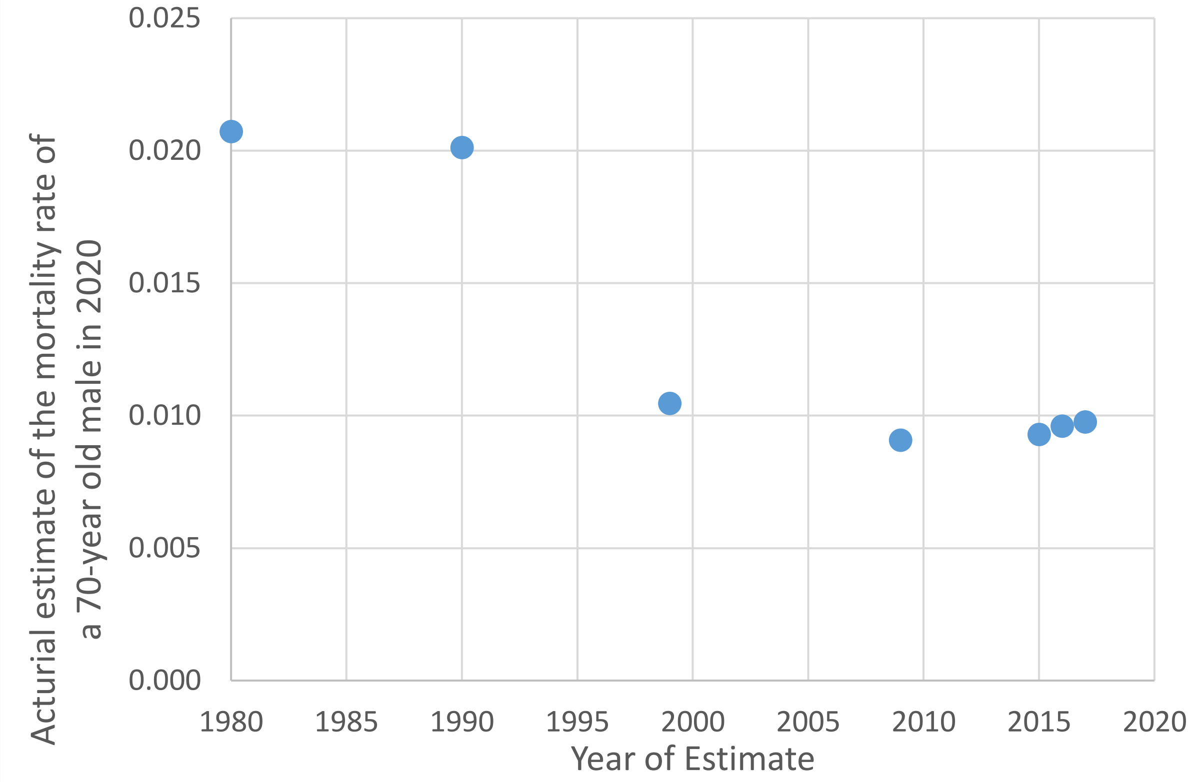
---

[325] See Browne and Owen (2019) for wider discussion of the history of actuarial mortality improvement projections.

[326] See Redington (1986), Section 5, for a selected collection of Redington's papers and speeches on mortality modelling.

[327] Redington (1986), p. 342.

Exhibit 5.1: Historical actuarial estimates[328] of the mortality rate of a 70-year old male in 2020



The mortality rate forecasts shown in Exhibit 5.1 are quite thought-provoking. They show that, although actuaries like to quote mortality rates to six significant figures, it is quite possible that the first significant figure of a 30-year mortality rate projection will be wrong. In the 1980s, actuaries (along with other professions and scientific communities with an interest in human mortality) strongly under-estimated the rate of improvement in the mortality rates of elderly people that would occur in the UK during the 1990s. And if records were readily available, they would likely show that mortality improvements for the 50-70 age group were significantly under-estimated during the 1970s and '80s. Going further back in history, the improvements in the mortality of infants would

[328] The data in this chart has been calculated from the mortality tables and mortality improvement formulae and spreadsheets provided by the Continuous Mortality Investigation (CMI). The 1980 estimate is derived from the base mortality rate for a 70 year-old male in a(90) (0.23412). This is the 1980 estimate of 1990 mortality. The mortality improvement factor from 1990 to 2020 is derived using the 'twentieths of age' rule described in the Preface of a(90) Tables for Annuitants (1979). That is, the mortality improvement factor uses the ratio of the mortality rate of a 68.5 year old to the mortality rate of a 70 year-old in a(90) (0.8854). The 1990 estimate is derived from the base mortality rate for a 70 year-old male in IM(80) (0.030309) and the 2020 mortality improvement factor (0.664) is derived from the formula presented in CMI Report 10, p. 52. The 1999 estimate is derived from the base mortality rate for a 70 year-old male in IML(92) (0.01838) and the 2020 mortality improvement factor (0.569) is derived from the formula presented in CMI Report 17, p. 93. The 2009 estimate is derived from the base mortality rate for a 70 year-old male in IML(00) (0.01625) and the 2020 mortality improvement factor (0.558) is derived from the CMI 2009 Mortality Projections Model. The 2015 estimate is derived from the base mortality rate for a 70 year-old male in LML(08) (0.012301) and the 2020 mortality improvement factors produced by the CMI 2015 Mortality Projections Model (0.755). The 2016 and 2017 estimates also use the base mortality rate for a 70 year-old male in LML (08). The 2016 estimate uses the 2020 mortality improvement factor generated by the CMI 2016 Mortality Projections Model (0.781) and the 2017 estimate uses the 2020 mortality improvement factor generated by the CMI 2017 Mortality Projections Model (0.794). The mortality improvement factor calculations from 2009 onwards require a long-term mortality improvement rate parameter assumption and 1.5% has been assumed in all cases here. The IFoA 2017 Mortality Bulletin states 1.5% is the most common parameter assumption used by actuaries (p. 20).

likely have been under-estimated during the second half of the 19th century and early part of the 20th century[329]. In recent years, an unexpected slowdown in mortality improvement rates has been observed across much of the developed world, and there is currently much debate and uncertainty around whether and how this slowdown should be extrapolated into the future.

To be clear, the above observations are not offered as grounds for criticising the quality of the modelling that has been used to produce the estimates plotted in Exhibit 5.1. It is not being suggested here that the modelling methods used in producing these estimates could have been readily improved in some obvious way. Rather, the argument offered here is that it is simply not possible to make reliably accurate long-term predictions of mortality rates. The drivers of long-term changes in longevity are the product of a non-stationary and inherently unpredictable social, cultural, political, economic, technological and physical environment. It is just too complex for reliably accurate quantitative prediction.

Consider the breadth of factors that may impact on long-term mortality improvements. Well-documented examples include: the outbreaks of new disease epidemics (consider HIV / AIDs in Africa in the early 2000s; the resurgence of old epidemic diseases and infections due to increases in antibiotic resistance; changes in lifestyle (obesity leading to type 2 diabetes, changes in smoking habits and alcohol consumption); political and economic factors (such as the State's commitment and ability to deliver a quality health care system, consider Russia in the 1990s); the levels of health spending[330] (which some studies have shown can be correlated with mortality rates[331]); income inequality and the potential divergence in longevity trends across socio-economic groups[332]; scientific and medical developments (consider how antibiotics vastly reduced deaths from infectious diseases in the 1950s; or how child mortality rates were significantly reduced by vaccination in the 19th century); and technological developments such as biometric monitoring and the 'Quantified Self' and how they may alter behaviours and lifestyles in ways that are consequential for mortality rates.

Much progress has been made in understanding what has driven historically observed mortality improvement patterns. It is now understood that much of the improvement in longevity that occurred in adults in the UK in the 1980s and 1990s was due to a reduction in deaths from cardiovascular disease (which, in turn, was partly the result of improved medical treatment techniques and technology and partly the result of a reduction in key behavioural risk factors such as smoking). Such analysis, whilst illuminating and useful for many purposes, still cannot be used to reliably predict how the propensity of various causes of death will change in the future. It is, of course, possible to say that the rate of change in cardiovascular disease as a cause of death cannot continue to fall at the rates it fell in the 1980s and 1990s, for the simple reason that it cannot cause a negative number of deaths. But beyond that trivial observation, little about the rates of different causes of death seems very predictable.

---

329 See, for example, Weiss et al (2019) for analysis of which age groups have contributed most to improvements in life expectancy during different historical periods.

330 There is a working hypothesis that the austerity policy of the UK government during the years following the global financial crisis, and the relative limits on social and health spending that were associated with those policies, are casual factors in the slower pace of mortality improvement that was observed in those years. See IFoA (2017), p. 15-17.

331 See Oliver (2019) for a fuller discussion and further references.

332 Between 2001 and 2015, English female mortality improved by 28% in the least deprived group but by only 11% in the most deprived group. See Browne and Owen (2019), p. 212.

The future evolution of the drivers of changes in human longevity, how they interact, and how they impact on different sections of society, are all highly uncertain and cannot be inferred from historical data or deduced from some easily identifiable causal structure. It might be objected that this point is well-recognised by the actuarial profession and always has been. After all, insurance companies have always held prudential capital in recognition of the risk that their long-term mortality projections could be materially wrong. Moreover, in recent decades, actuaries have developed stochastic models of mortality risk. The first was published in 1992[333]. Since then, the number of such models has proliferated, offering model structures of increasing complexity, and producing a range of alternative probability distributions for the future evolution of mortality rates. This is a fair and reasonable argument. Nonetheless, it is contended here that a clearer recognition of the significant epistemic limits to the reliable accuracy that is possible for long-term mortality projections can have major implications for the way actuaries build and use actuarial models, and how mortality-related risks are managed by the institutions actuaries serve. These implications are explored further in Chapter 6.

Actuarial Models and Long-term Predictive Success? Case Study (2): Interest rate projections

We next turn to a completely different type of phenomenon for which actuarial science produces and makes use of long-term predictions: the long-term interest rate.

Actuaries use long-term forecasts of the behaviour of the long-term interest rate for a range of purposes in various fields. The analysis of the advance funding of defined benefit pension funds is a significant example. There, actuaries use their projections of the behaviour of the long-term interest rate to estimate the pension fund's required long-term contribution rate; and to estimate the amount of assets required today to fund liability cashflows as they fall due; as well as to advise on related topics such as the pension fund's investment strategy. In life assurance, long-term projections of long-term interest rates can be used to analyse the run-off solvency of the business, and to analyse the long-term outcomes related to investment and risk management strategies.

Before the 1980s, actuaries tended to use only deterministic projections of these interest rates. The recommendations of the Maturity Guarantee Working Party[334] introduced the stochastic modelling of equity returns into actuarial practice (specifically, for the purposes of reserving for long-term maturity guarantees in unit-linked life policies). And David Wilkie, one of the prominent members of that working party, further developed the stochastic modelling of equity returns into a broader stochastic asset model that also included the modelling of interest rates and inflation (and, later, property). This broader model became known as the Wilkie model and was widely used by UK actuaries in life and pensions throughout the late 1980s, 1990s and, perhaps to a slightly lesser degree, well into the 21st century.

Wilkie's paper presenting this long-term stochastic asset modelling was published in 1984[335]. This research represents the earliest formal stochastic actuarial model of the behaviour of the long-term interest rate. The model was calibrated to historical data from 1919 to 1982. This may sound like a long historical period and a lot of data. But bear in mind that the model is explicitly intended for *long-term projection*[336]. What do we mean by long-term? In actuarial terms, this usually means somewhere between 10 and 40 years. The calibration data provides a sample of only 7 non-overlapping 10-year observations, and even less for longer horizons. The UK long-term interest rate

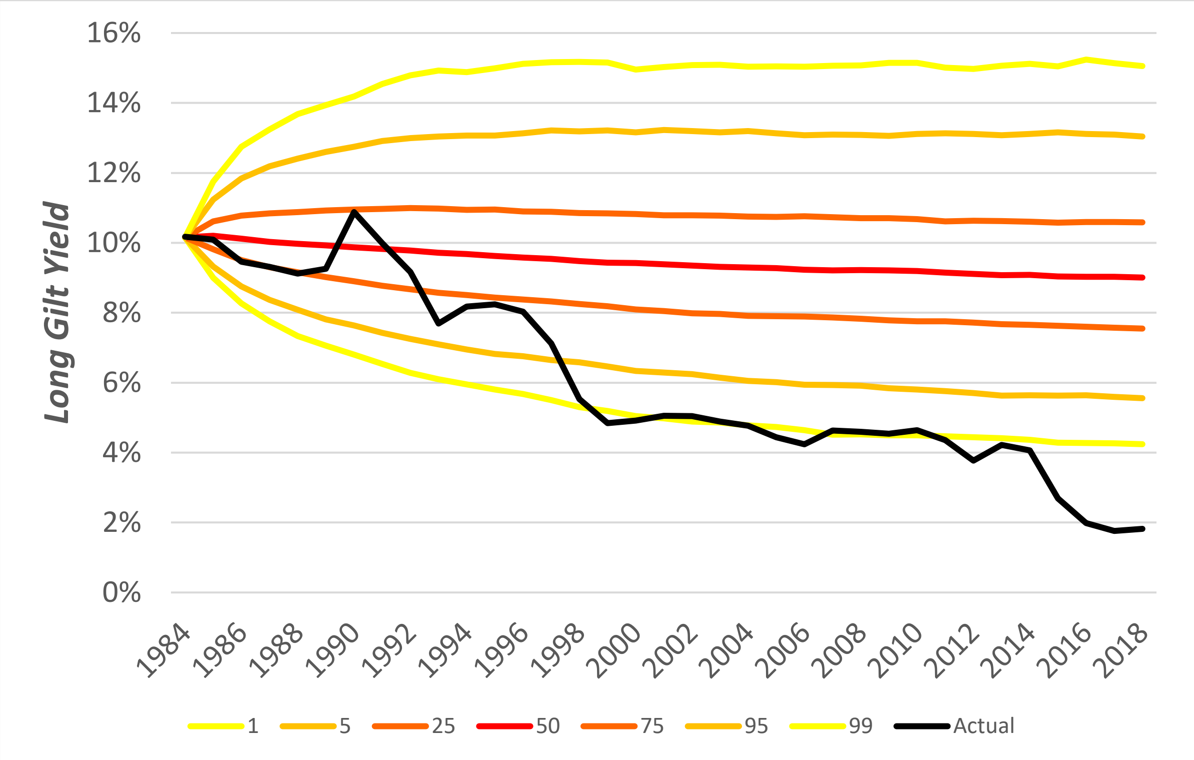
---

[333] Carter and Lee (1992)
[334] Ford et al (1980)
[335] Wilkie (1984)
[336] See Wilkie (1984), Section 1.3.

varied between less than 3% and almost 12% between 1919 and 1982[337]. Even in a statistically perfectly behaved stationary environment – Knight's world of measurable risks - a sample of 7 can provide quite limited information about the phenomenon's average level, given this degree of variability. And it can provide much less information about the behaviour in the tails of the probability distribution (again, even before considering the potential difficulties of immeasurable uncertainty that may arise from non-stationarity).

So, before even turning on our computers, it is clear that Wilkie set himself an ambitious objective when he set out to develop a model that could reliably estimate the multi-decade tails of the long-term interest rate. Let us now examine what the model projected. Exhibit 5.2 shows the percentiles of the distributions of the future long-term interest rate that were produced by the model in 1984. The chart also plots the actual long-term interest rate that occurred over this period[338].

Exhibit 5.2: Percentiles of the projected long-term interest rate distribution from Wilkie (1984) model, and the actual long-term interest rate path



The UK long gilt yield was 10% in 1984. The Wilkie model generated a long-term median for the projected long-term interest rate at a similar, if slightly lower, level – the model produced a median value for the 2018 long rate of 9%. This was consistent with the model's assumption of an unconditional average inflation rate of 5% and an unconditional average real interest rate of 3.5%. According to the model, there was a probability of 1% that the long gilt yield in 2018 would be greater than 15%, and a 1% probability that it would be less than 4%. The actual 2018 outcome for the long-term interest rate was 1.8%. This was a deep outlier in the context of the projected distribution - an outcome this low was a 1-in-10,000 event according to the model. In summary, **the**

[337] Data freely available on Bank of England website.

[338] In the Wilkie (1984) model, the long-term interest rate is represented by the modelled consol yield. For the actual long-term interest rate, the consol yield has been used from 1984 to 2016; and the 30-year gilt yield has been used for 2017 and 2018. Historical consol yield data is freely available from the Bank of England website.

1984 model attached virtually zero probability to the long-term interest rate of 2018 being as low as the actual 2018 long-term rate.

Wilkie presented an updated version of his stochastic asset model in 1995[339]. This update re-calibrated the model with the dataset extended to 1994. Wilkie also decided that this time he would exclude some of the early years of the dataset because they were in the close aftermath of the First World War, which he now judged was an exceptional period that should be ignored by the calibration process. So, the updated model was calibrated to the period 1923 to 1994.

The 1995 update also included a model for the index-linked gilt yield[340]. Index-linked gilts have been issued by the UK government since 1981. At the time of the 1984 paper, there was insufficient data to calibrate a model. But Wilkie judged that enough experience had been observed by 1995 to produce a long-term stochastic model of index-linked yields. Exhibit 5.3 shows the percentiles of the distributions of the future index-linked yield that were produced by the model and its calibration in 1995. The chart also plots the actual long-term index-linked yield that occurred over this period.[341]

Exhibit 5.3: Percentiles of the projected index-linked gilt yield distribution from Wilkie (1995) model, and the actual index-linked yield path

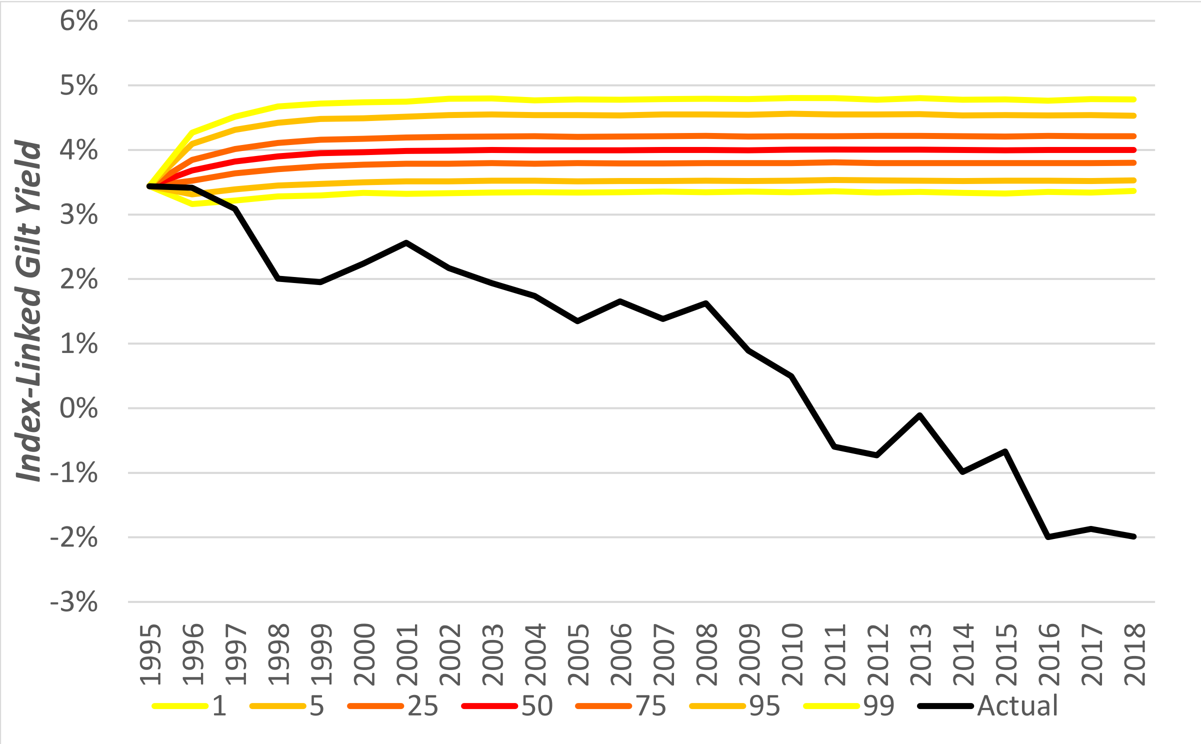


Exhibit 5.3 provides a good illustration of the perils of extrapolating historical socio-economic data for the purposes of making long-term probabilistic statements.

The above analysis suggests the Wilkie model(s) failed to attach meaningful probabilities to the long-term interest rate outcomes, for both nominal and index-linked rates, that have been realised since the publication of the model. There are (at least) three alternative ways of interpreting these findings:

[339] Wilkie (1995)

[340] Wilkie (1995), Section 9.

[341] Historical index-linked yield data is freely available from the Bank of England website.

1. It could be argued that the models delivered something closely approximating the true probability distributions for the future behaviour of the long interest rate, and the true probability distributions for the future index-linked yield, and something extremely unlikely occurred in both cases.
2. Alternatively, it could be argued that the probabilities of interest rates as low as those that actually occurred occurring were much higher than the probabilities produced by the models. It might further be argued that a better model and calibration could have been developed that would have performed much better in these respects.
3. Or we might assert that the nature of interest rate behaviour creates a natural epistemic limit that means it is simply is not possible to develop long-term models for interest rates that can make reliably accurate probability statements about the range of outcomes that that will occur.

Let's take each of these three alternative arguments in turn, starting with the first one. As noted in Chapter 2's discussion of the falsification of probability statements, it is generally not possible to irrefutably prove that statements of the type used in the first argument are wrong – there is always a chance that the model was correct to attach a very low probability to the given event happening, and the very low probability event did indeed occur. In the particular instance of the long-term outcome for the long-term interest rate, it requires a highly unlikely event to have occurred – one with a 1-in-10,000 probability. It is a possible explanation, and if we can convince ourselves that the probabilities that should be attached to answers 2 and 3 are lower, then we could conclude it is the most likely explanation (an example of a Bayesian probability logic that does not attempt to quantify the probabilities attached to these three statements!).

But in the case of the index-linked yield, this 'non-falsifiable' logic does not apply. That is because the model assumes that index-linked bond yields are lognormally distributed. This means the model says that it is *impossible* for the index-linked yield to be less than zero. As can be seen in Exhibit 5.3, index-linked gilt yields have been less than zero since 2011. As a result, it can be concluded with certainty that the index-linked yield model did not produce an accurate representation of the true long-term probability distribution of the index-linked yield. So, we conclude that the first argument is a deeply unsatisfactory explanation of what is observed in exhibits 5.2 and 5.3.

Let's move on to the second argument. The Wilkie (1984) model has received some significant criticism from other actuaries and statisticians with respect to the statistical methodology that was employed in the choice of model structure and its calibration[342]. And we noted above that Wilkie himself revised the models and their calibrations in 1995. This resulted in some material changes in the projected probability distributions of the inflation rate and the consol yield[343]. Furthermore, there is an array of subjective modelling and calibration choices that could have been made in 1995 that would have produced different probability distributions for the future levels of the index-linked yield. Some of these modelling approaches may be more sophisticated than those used by Wilkie. However, it is contended here that no conventional method of modelling and calibrating a probability model for the long-term projection of the index-linked gilt yield in 1995 would have attached a meaningful probability to the yield being as low as -2% in 2018. *And if the model did*

[342] See, for example, Geoghegan et al (1992); and Huber (1996), Section 8 and 8.4.1 in particular regarding the consol yield model and calibration.
[343] Wilkie introduced a more sophisticated model of inflation, with time-varying volatility. These changes increased the variability of the projected consol yield.

suggest such a scenario had a meaningful probability, it would almost certainly have been rejected with incredulity by experts at the time.

It might be argued that actuaries would have rightly rejected such a model as silly because such an outcome really was extremely unlikely, and so we are back to argument 1. It is only extremely unlikely, however, if it is assumed that the socio-economic environment is statistically stable and quantifiably measurable. And all of the model and calibration enhancements that may be described under argument 2 have to subscribe to such an assumption. Wilkie's described the 1995 ARCH extension of his inflation model as a way of acknowledging that inflation volatility is non-stationary. But really, such a model extension simply describes a slightly more complicated form of stationarity, where all the uncertainty in the behaviour of the phenomena still remains quantifiable. Re-quoting Redington one more time: 'If we cannot foresee the future then we cannot! No proliferation of polysyllables or multivariate analysis will succeed (other than in deceiving ourselves).'

The failure to deliver reliably accurate predictive results is not a consequence of mere sampling variation (argument 1), nor of a failure of technical modelling skill (argument 2). It is a consequence of the use of a fundamental premise that does not hold in our socio-economic world – the assumption of a form of stationarity in the behaviour of empirical phenomena such as interest rates and mortality rates. The socio-economic environment is not stationary. It is constantly changing in important ways and its implications for inflation, interest rates and most other socio-economic phenomena are complex and inherently unpredictable. It always been this way. *The only thing we can reliably predict is that our socio-economic system will continue to change in unpredictable ways*. Once we accept that, it becomes clear that these modelling approaches cannot be expected to produce reliably accurate probability distributions.

This perspective logically leads to the third of the arguments considered above. The above anecdotal analysis, whether of mortality rates or interest rates, strongly supports the case for this argument. The fundamental complexity of these phenomena creates a natural and intuitive epistemic limit that implies it is simply not possible to make reliably accurate long-term probabilistic predictions for the behaviour of these types of socio-economic phenomena. It is a simple argument with a clear philosophical basis that has attracted support from many important thinkers over the last 200 years or so (we have already mentioned above a diverse group of examples such as Venn, Keynes, Hayek, Popper, Knight and Redington). If we accept this perspective, the key question that remains is: what are the implications of this conclusion for the methodology of actuarial science? *What does an actuarial science that accepts this argument look like?* This is the main topic of Chapter 6. But we are not quite finished with our analysis of current methodology.

5.5 Model risk and uncertainty

The reflective actuary, upon reading Chapter 5.4, might reasonably retort that much of what it says is obvious. Every actuary knows the old aphorism that models are always wrong and sometimes useful. Actuaries are very aware of the limitations of models and work hard to manage its consequences in sophisticated ways. Methods to measure the extent to which models may be wrong and to mitigate the consequences of these errors come under the banner of 'model risk management'. A lot has been written in recent years by academics, actuaries and financial regulators about how to measure and manage model risk. This section reviews some contemporary thinking on model risk and its recommendations for good model risk management. These arguments are considered in the context of the empirical analysis of Chapter 5.4 and the philosophical perspectives that have been used to explain it.

Post-Global Financial Crisis perspectives on model risk from the US Federal Reserve, UK HM Treasury and Bank of England

Interest in the topic of model risk management was heightened by the Global Financial Crisis of 2008. It is well known that risk models of financial institutions, especially banks, did not have a good crisis. Commercial and investment banks use quantitative financial models as extensively as insurance groups do – in pricing, underwriting, valuation, risk and capital assessment and risk management. Some banks' risk models did not perform well through the Global Financial Crisis. In 2008, the Goldman Sachs Chief Financial Officer famously commented that the firm's internal risk models experienced several 25-standard deviation one-day events over the course of a couple of weeks.

A lot has been written about model risk in the years since the Crisis, and we will review a handful of important papers. Some of these pieces have been very influential in the field of model risk in finance in general, and some are specifically germane to actuaries and actuarial science. We will first consider two important papers produced by government institutions. In an influential piece of supervisory guidance published in 2011[344], the US Federal Reserve defined model risk as 'the potential for adverse consequences from decisions based on incorrect or misused model outputs and reports'[345]. A couple of years later, HM Treasury published a review of the quality assurance that was applied to the models in use across the UK government and its civil service[346]. Whilst both these papers were written for what might be regarded as predominantly non-actuarial institutions (the Fed paper was written for the banks that are regulated in the Federal Reserve System; the HM Treasury paper was written for government and civil service departments), they both provide useful and intelligent guidance on how to manage the model risk that arises from the use of models in financial institutions, including actuarial institutions such as insurance firms and pension funds.

The Federal Reserve guidance noted that a model is necessarily a simplification of reality. This limits the applicability of the model. It means that a model that may be fit for one purpose may not be fit for another. Similarly, a model that is fit for purpose at some given point in time may not be fit for purpose at another time. The Treasury review similarly noted that models can provide 'insights and understanding, only if they accurately reflect the policy environment and are used correctly'.

The US and UK guidance emphasised the importance of 'effective challenge' and 'critical review', and a culture (British) and set of incentives (American) that supports it. Effective challenge may come in the form of external validation exercises and peer review. The papers recognise that model risk can arise from operational implementation errors such as mistakes in computer code or misinterpretations of data. The regulatory prescription for the management of these forms of operational errors is largely based on common sense: good practice should include extensive documentation of model development, calibration and validation; there should be effective control environments and checklists; annual model reviews should take place; this may include internal and / or external peer review of the model; there should be clear ownership of model development; policies and procedures to govern and oversee these activities should be in place; model inventories should be produced and kept up to date; and so on.

The US Fed guidance also implores banks to ensure the model is 'statistically correct'. The note acknowledges, however, the logical limitations that must arise in establishing such correctness: 'appropriate statistical tests depend on specific distributional assumptions….in many cases statistical

---

[344] Board of Governors of the Federal Reserve System (2011)
[345] Board of Governors of the Federal Reserve System (2011), p. 3.
[346] HM Treasury (2013).

tests cannot unambiguously reject false hypotheses or accept true ones based on sample information'[347].

Moving beyond operational errors, the US Fed guidance also recognises that there is another, more fundamental, source of potential error in models: 'Because they are by definition imperfect representations of reality, all models have some degree of uncertainty and inaccuracy'[348]. The guidance points to intuitive ways of quantifying this uncertainty: consider the confidence intervals of the model's parameter estimates and the model's results; try to quantify the potential impact of factors not included in the model.

The guidance also recognises that quantitative estimation of model uncertainty may not always be possible. Or, put another way, it recognises that the usual approach to quantification of uncertainty is reliant on assumptions that are untestable and may be intuitively unlikely in many situations. To understand this fundamental point in practical terms, let us re-consider the Wilkie (1995) index-linked yield model discussed in Chapter 5.4. If it is assumed that the specified model structure is the true and permanent stochastic process for the future behaviour of index-linked yields and that all the historical data used in the calibration has also been generated by this same process, it is quite straightforward to estimate the parameter uncertainty associated with the model calibration. Indeed, Wilkie does this, showing that his parameter estimates have assuredly low standard errors[349].

But these estimated errors are only low because all of the most important sources of error have been assumed away in the estimation process! We have no reliable justification for the assumption that the posited model structure is the 'true' model. There is no testable scientific theory that has been used to determine that this model is the true description of the future behaviour of real interest rates (and Chapter 4's discussion of the methodology of economics and econometrics highlighted that none is likely to be forthcoming). It is just one model choice from a choice of many, many possible choices that could be said to be similarly consistent with the data. The model choice can be viewed as the most parsimonious structure that was consistent with a small number of key premises such as: real interest rates cannot ever be negative; real interest rates have a constant and known unconditional mean; real interest rates have a constant rate of mean-reversion to this mean level. Wilkie's model choice does an excellent job of being consistent with those premises and the available historical data. But the premises are deeply inconsistent with empirical reality. The model will therefore be unreliable as a probabilistic description of future real interest rates, and this will be the case irrespective of how small the standard errors produced by the calibration process are.

It is important to note, again, that this is not a particular criticism of Wilkie's technical methods. Wilkie's methods were quite within the domain of orthodox statistical model-fitting techniques. In Huber's often-critical review of Wilkie's asset modelling, he argued that the small data sample meant that the real interest rate model 'should be used with caution in long-term studies'. But even he concluded that 'this model appears to be satisfactory'[350]. The fundamental point here is not that a different statistical fitting method could have done a better job than another – it is that all these methods must rely on assumptions that have no reliable basis in the context of economic phenomena like interest rates and inflation rates.

[347] Board of Governors of the Federal Reserve System (2011), p. 6.
[348] Board of Governors of the Federal Reserve System (2011), p. 8.
[349] Wilkie (1995), Table 9.1, p. 883.
[350] Huber (1996), p. 242.

The main mitigation strategy advocated by the US Fed guidance in the presence of unquantifiable uncertainty is to add prudent margins to model parameters or the model output 'in the interest of conservatism'[351]. This is clearly meant as a pragmatic supervisory solution, but its methodological logic is not very satisfactory. Would increasing the volatility parameter of the Wilkie index-linked yield model by a margin of say, 50%, have ensured a 'conservative' description of future real interest rates? *No.* Would using the 99th percentile as a conservative estimate of the 90th percentile have ensured a prudent measure of the 2018 90th percentile? *No.* Once the presence of this type of uncertainty is recognised, assessing arbitrarily 'conservative' probability estimates is just as impossible as assessing accurate estimates. In the presence of this form of unquantifiable uncertainty, the basic objective of generating accurate quantitative measurements of probabilities is a task fundamentally beyond the model.

The US Fed guidance defines effective model validation as 'an assessment of the reliability of a given model, based on its underlying assumptions, theory, and methods', and is strongly recommended[352]. Annual validation reviews that compare actual outcomes with the model's probability predictions can be used to identify when the model should be 'redeveloped'. This is unequivocally a good idea. Again, referring to the Wilkie real interest rate model example, this would probably have resulted in the model being identified as no longer fit for purpose by 1998. But *how* is the model to be 'redeveloped'? Having been proven fallible once, what is the magic fix that will provide the model with reliable accuracy going forward? There isn't one. When and how does this effective model validation process reach the obvious and inescapable conclusion: these types of model failures are not merely a result of process errors or poor modelling decisions; there are no magic modelling fixes that can suddenly be unearthed with the aid of one more year of data; and it is counter-productive to labour under the illusion that this is the case.

We now turn to a third important post-GFC paper by supervisory authorities on model risk and uncertainty. The then-Chief Economist at the Bank of England, Andrew Haldane, co-authored a notable paper[353] that argued that model uncertainty was an inevitable feature of financial risk modelling and that this had profound implications for the use of models in risk and capital assessment. Model uncertainty, it was argued, placed a 'heavy reliance' on professional judgement and experience[354]. So far, this is consistent with the findings developed above. However, rather than arguing, as has been argued above, that this inevitable model uncertainty means that positivist modelling is doomed to inevitable failure, the central argument of the paper was that the presence of inevitable model uncertainty means that, for the positivist application of generating a probability-based capital assessment answer, much simpler models are preferable to the complex internal models commonly developed by banks.

The general logic of this argument, first developed by the Nobel Prize-winning economist Herbert Simon in the 1950s, is that simpler methods are more robust to the lack of knowledge implied by model uncertainty (we can loosely think of 'robust' here as meaning the answer is less sensitive to being wrong about things that are uncertain; Simon argued that humans operating in a world of uncertainty use simple rules or 'heuristics' to survive and prosper[355]).

---

[351] Board of Governors of the Federal Reserve System (2011), p. 8.
[352] Board of Governors of the Federal Reserve System (2011), p. 10.
[353] Haldane and Madouros (2012)
[354] Haldane and Madouros (2012), p. 117.
[355] Simon (1956)

Haldane's paper also argued that illustrative simple methods (such as leverage ratios) empirically outperformed the complex internal models used by banks in predicting bank failure during the crisis. Mervyn King, the former Bank of England governor, made a very similar argument in his 2016 book, *The End of Alchemy*[356]. And we noted earlier that Redington had similarly argued in favour of the use of simple models in the presence of model uncertainty.

Simplicity is undoubtedly a modelling virtue…but simplicity is not a comprehensive solution to the inevitable failings of positivist modelling in a world of immeasurable uncertainty. Once the presence of material model uncertainty is accepted, it is difficult to find virtue in any mechanical model-based probabilistic capital assessment, simple or complex. In such circumstances, there is the risk of systematically under-stating future probability tails as a result of assuming away sources of uncertainty without justification. As argued above, many different model structures may provide a set of historical data with excellent parameter calibration performance, under the assumption that the given model is correct and there is no model uncertainty. Simplifying the model structure does not solve these problems that arise in a positivist use of the model as an answer-generating machine.

So, simplicity is not, in and of itself, the solution. Moreover, model complexity can be valuable, when used in the right hands. When models are used as interpretivist tools rather than positivist answer-generators, complexity is capable of producing valuable insights (for experts that understand the models) that simple models simply cannot. This suggests we should be reticent to conclude that complex models have no role to play, especially in the context of some of the complex balance sheets that have been created by financial institutions (whether or not *that* complexity is a good idea is a different question, and one that is discussed in Chapter 6.1).

There is another feasible explanation for why Haldane found that simple measures could outperform the complex internal models of banks in the presence of uncertainty: the complexity of the banks' models was never actually intended to provide a more accurate measure of risk and capital. Rather, the complex models resulted (at least in part) from banks' incentives to find ways of producing lower measures of risk and capital requirements, and regulators' failure to effectively supervise how banks' developed and used these internal models.

In conclusion, complexity may well be at best pointless in a positivist form of model use. Above we have argued that this is not necessarily a strong argument for more modelling simplicity, but an argument for moving away from the positivist form of model use. As noted above, the US Fed paper started by defining model risk in the context of the decisions made with the use of a model. This is the most fundamental point, and it points to a better resolution of the above difficulties. As argued elsewhere above and below, the presence of unquantifiable uncertainty does not render a quantitative probabilistic model useless. Far from it. Instead, it changes how the model should be used.

The model's purpose should not be to mechanically generate numbers that are interpreted literally by laypeople who have been assured of the modelling process's robustness and objectivity by experts' rubber stamps. Rather, the use of a model in the presence of unquantifiable uncertainty should be as an interpretative tool for experts to use to gain new insights that are useful for their professional output.

[356] King (2016)

Recent Actuarial Literature on Model Risk and Uncertainty

In recent years, the UK actuarial profession has also produced some notable research on the topic of model risk and uncertainty. A Model Risk Working Party published two papers on the topic in the late 2010s[357]. These papers explicitly considered the idea of model uncertainty discussed in the US Federal Reserve and Haldane papers. The Introduction to the first of these actuarial papers notes:

"The randomness of variables such as asset returns and claim severities makes estimates of their statistical behaviour uncertain: in other words, we can never be confident that the right distribution or parameters have been chosen."[358]

The paper explicitly recognises the unquantifiable nature of model uncertainty. Or, in the paper's terminology, how the assessment of model uncertainty also inevitably involves measurement uncertainty:

"A quantitative analysis of model uncertainty, e.g. via arguments based on statistical theory, will itself be subject to model error, albeit of a different kind."[359]

Like the US Federal Reserve and HM Treasury papers, these two UK actuarial papers advocate a model governance framework that includes clear model ownership, senior organisational oversight and a policy that sets out minimum standards for model quality assurance that includes validation and appropriate documentation. There are also some additional suggestions. For example, the 2016 paper suggests that the Board of a financial institution that uses models should define its 'model risk appetite', which the paper defines as the 'extent of its [the Board's] willingness, or otherwise, to accept results from complex models'[360].

The phrase 'to accept results' suggests the Working Party envisages models being used in the specifically positivistic sense of generating answers for lay authorities (in this case the Board of the financial institution), rather than as tools to provide the institution's experts with insights that they can use in forming their advice to the lay authority. It also implicitly seems to suggest that model risk is only present in complex models. As was argued above, model uncertainty is not mitigated by simplicity: a very simple model of the expected future real interest rate will have unavoidable and material model uncertainty.

The 2016 paper suggests that model documentation should include 'a reasonable range or confidence interval around the model result (e.g. by using sensitivities to key expert judgements)'[361]. As noted above, the paper had already pointed out that estimates of model uncertainty would be inherently subject to error, and so the reference to expert judgement is quite natural. The paper expands on this point more explicitly later:

"An overlay of expert judgment should be applied to the model output to address the uncertainty in the model. For example, if there is significant uncertainty in one of the underlying assumptions and hence there is a range of plausible results, then an expert judgement may be applied to identify a more appropriate result within the range of reasonable outcomes. Alternatively, the expert

[357] Morjaria et al (2016); Black et al (2018).
[358] Morjaria et al (2016), Section 1.3.
[359] Morjaria et al (2016), Section 1.3.
[360] Morjaria et al (2016), Section 2.4.
[361] Morjaria et al (2016), Section 2.8.

judgement may be an entirely independent and objective scenario assessment to complement the modelled result, or replace the use of the model altogether."[362]

There is much sense in this advice. For example, independently produced and specific 'stress scenarios' (for example, the effect on longevity improvements of a major breakthrough in a form of cancer treatment) can help to externally validate whether the probability tails produced by a mortality projection model are of a sensible order of magnitude. But it should be borne in mind that once the inherent uncertainty in the future behaviour of a phenomena is recognised, this uncertainty is attached to the development and interpretation of specific stress test scenarios for the phenomena as well as to its estimated probability distributions. And the identification of such stress test projections may be subject to similar biases and errors in the underappreciation of the extent to which socio-economic systems can diverge from the norms of recent past. For example, an expert economist, when asked in 1995 to produce some independent stress test scenarios for the long-term real interest rate in 2020 would have been unlikely to have identified many plausible circumstances in which a rate of -2% could arise.

Although the paper acknowledges that model uncertainty is inherently difficult to quantify, it nonetheless suggests that approaches such as those suggested above should be used to determine if the firm's use of models exceeds the Board's specified model risk appetite. The paper argues that it is up to the Board to decide whether it is willing to accept the use of a model that comes with a given estimated level of model uncertainty. The paper implies this should be a binary decision: either the model has too much uncertainty and the Board should reject its use; or the model has an acceptable amount of model uncertainty and should therefore use its results.

To re-iterate the point made earlier, the perspective developed in this work is that it is possible for a model with a large amount of model uncertainty to be put to good actuarial use. Not, however, for the purpose of mechanically generating results for the Board (even if accompanied with the assurance of experts that the model has acceptably low model uncertainty), but to provide understanding and insight to the firm's actuaries and other technical experts (who can derive useful insight from interpreting the model's outputs in the presence of uncertainty, and then use this insight to provide professional advice to the Board that has been enhanced by this interpretative expert use of the model).

In one particular sense (but not in others, as will be discussed in Chapter 5.6 below), this argument suggests a shift back in time to how actuaries, and perhaps other financial risk professionals, used models prior to the advent of regulatory principle-based capital assessment and the permitted use of internal models (which began in banks in 1996 with the Market Risk Amendment to Basel I; was then extended to include the assessment of credit risk in Basel II in 2004; was introduced into UK insurance to some degree by the Individual Capital Assessment system in 2003; and plays an important role in Solvency II, which was introduced to EU insurance supervision in 2016. Actuaries' mechanical use of model output to assess risk-based capital requirements can be traced a bit further back to the pioneering work of Sidney Benjamin and the Maturity Guarantees Working Party that followed).

These regulatory approaches provide firms with the opportunity to develop their own models for risk-sensitive probability-based capital assessment. One of the fundamental ideas that underlie the contemporary approach to internal model-assessed capital requirements is that sophisticated

[362] Morjaria et al (2016), Section 2.9. The paper also discusses an example of the application of this approach in the context of longevity models in Section 4.3.

scientific models can be developed and validated by experts such that relative laypeople can be assured that the model will then reliably produce the right answer. The process requires that the firm's Board and the regulatory authorities are, after a rigorous model validation process, convinced that the experts have done their work adequately in creating the answer-generating model, and the model is then put to work to determine the required amount of capital.

The interpretivist perspective implies a different sequencing. Instead of experts building models that generate answers for Boards and regulators, experts can *use* models to help *them* provide answers to Boards and regulators. This may sound like a nuance, and this distinction may not be quite so black and white in real life. But it has some fundamental implications for the way some actuarial models are used today. It removes the unrealistic idea of Boards of modelling laypeople taking responsibility for complex model outputs and their use. It empowers the expert to use their expertise to solve problems rather than to merely build models. But the obvious shortcoming of such a system is it places reliance on the individual expert's subjective judgment. If the expertise cannot be distilled into a scientific model, how can the expertise be trusted and demonstrated to be 'correct'?

Transparency is the most obvious answer. One natural solution would be to require the experts' capital assessment and the expert's reasoning to be made publicly available to other experts. This is reminiscent of the 'freedom with publicity' system of many years ago. In this setting, the expert actuary can essentially use whatever methods and reasoning he or she sees fit to assess the capital requirement, providing those methods are disclosed in sufficient detail for other experts to review and opine on the reasonableness or otherwise of their methods. This is likely to result in better reasoning and methods than the current process of internal model governance where a hard-stretched regulator is given the task of privately reviewing each insurance firm's complex model in order to be convinced that the model can meet their objective of answering unanswerable questions. As noted above, Haldane's empirical analysis suggests this approach may be resulting in internal capital models with serious inadequacies. Would the same failings be as likely if the models, calibration and assumptions were fully exposed to the harsh sunlight of public disclosure and peer group scrutiny?

5.6 Economic theory and actuarial science

The application of economic theory in the context of actuarial science has been a topic that has generated considerable debate within the actuarial profession over the last 30 or 40 years. In Pemberton's 1999 methodology paper and its discussion, both Pemberton and Cartwright were strongly sceptical of the usefulness of economic theory to actuarial science. To quote Cartwright:

"We should be wary of principles and methods imported from theoretical economics which are, themselves, often borrowed from modelling strategies used in physics, where phenomena tend to be regular and repeatable, and predictions are, for the most part, confined to the well-controlled circumstances of a laboratory."[363]

The abstract nature of economics, and its empirical failings as a positive science as discussed in Chapter 4, mean that actuaries should indeed be highly sceptical of, say, the quantitative predictions of econometric models for actuarially relevant economic variables such as inflation or salary growth. But does this mean that actuaries should have no use for economic theory? The perspectives developed in this work suggest the more interesting question is: *can actuaries make use of economic theory in an interpretivist rather than positive sense*? Chapter 4 concluded that it is the case

[363] Cartwright in Pemberton (1999), p. 178.

generally that economic theory has real value as an interpretivist field of study. It will be similarly argued here that it is also the case that the interpretivist application of economic theory can provide profound insights in many areas within actuaries' professional domain.

The value of economic theory to contemporary actuarial science is potentially very substantial because, for one reason or another, many actuarial financial institutions have opted to substantially increase the amount of financial market risk on their balance sheets over the last fifty years or so (Chapter 6 will discuss whether this risk-taking is desirable, but the current discussion merely requires us to recognise it is present).

There are a handful of key ideas in financial economic theory that can provide useful insight into the costs of bearing those risks and they can be managed[364]. Several of these ideas give insight into what, in well-functioning markets, should *not* matter. For example, diversifiable risk should not be rewarded; the risk premia of underlying assets should not affect the price of options on those assets; the capital structure of a firm should not affect its total value; investors should not be expected to be able to outperform the market on a risk-adjusted basis after costs. And there are also ideas that are important formalisations of intuitive knowledge: for example, options can be thought of as continuously rebalanced dynamic portfolios of the option's underlying asset and the risk-free asset.

All of these insights result from theoretical deductions from unrealistic premises. They are the result of severe abstractions and the results should not be interpreted over-literally. But the logic of their derivation is sophisticated and can be highly enlightening. A deep understanding of these ideas can help to cultivate an economic logic that can deliver empirical insights for financial risk management and decision-making. Option pricing theory is probably the most important of these ideas to actuaries and can provide the best examples of how economic theory can be of practical use to the actuary. Conceptually, forms of financial options pop up all over actuarial science: in asset pricing (all debt can be thought of as having an option-based pay-off); in life assurance liability valuation (various forms of policy guarantee can be expressed as options); in the risk management of those liabilities (based on the insight that they have embedded options); in Defined Benefit pension fund sponsor default risk; and so on. The economic logic of dynamic replication can provide profound insight into the fundamental nature of a financial risk - its scale and sensitivities and the cost of transferring it to a third-party.

As an example for illustration, let's consider the somewhat infamous topic of the management of Guaranteed Annuity Options (GAOs) in the UK in the 1990s and the potential application of option pricing theory to this problem[365]. The GAO crisis was a complex problem, especially in the context of with-profit business, where legal interpretations of policyholders' reasonable expectations had a significant impact on who should and would bear the costs of the GAO liabilities. It is therefore not being suggested here that the derivative pricing theories of financial economics would have necessarily been a panacea for the multi-faceted GAO crisis (that had been built over several decades).

Nonetheless, there is a strong argument that UK life actuaries in the early-1990s could have found much insight into their emerging GAO issues from the then well-established financial economic theory of option pricing[366]. In particular, actuaries could have familiarised themselves with ideas

---

[364] See Turnbull (2017), Chapter 4 for an overview of the history of financial economics that is written for a mainly actuarial audience.
[365] See Turnbull (2017), p. 222-231 for further discussion of this historical episode.
[366] Boyle and Hardy (2003) also argues that a deeper actuarial understanding of financial economics would have helped to manage the GAO problem.

such as: the theoretical concept of an arbitrage-free price for the GAO and how an indicative value of that price could be estimated; how that price differed from the actuarial reserves then held for the GAO; the sensitivity of that price to future interest rate movements; how the arbitrage-free price implied a replicating portfolio for the GAO pay-off and how the composition of that replicating portfolio radically differed from the way in which GAO reserves were invested. These concepts could have provided actuaries with real insight into the nature, potential scale and route to risk management of GAOs. Using option pricing theory in this way does not mean subscribing to the belief that the number produced by the derivative formula is 'correct', or that the theory's assumptions are true. It is simply another tool that can significantly illuminate and provide a deeper, richer understanding of the financial problem at hand.

The actuarial use of the models of financial economics as interpretivist tools is not a new idea. In his doctoral thesis on the use of economic models in actuarial science, Huber wrote[367]:

"Although these theoretical models [based on financial economic theory] are clearly unrealistic, they have important pedagogic value. They provide decision makers with information about an understandable idealised environment. This establishes a basis from which they can make their own subjective adjustments if they wish. This illustrates the importance of being able to interpret actuarial economic models…*These models do not produce final answers for applications; they merely assist actuaries in understanding the relevant issues*." [my italics]

The above quotation is fully consistent with the perspective developed over Chapters 4 and 5 of this work. However, an important caveat is required to the argument that an interpretative use of economic theory can be a powerful asset to actuaries. The important practical insights offered by option pricing theory require a relatively deep understanding of the theory, its assumptions and nuances. A superficial, or flawed, understanding of the theory may lead to misinterpretations that make the application of the theory counter-productive. This raises interesting and important questions around professional training and the development of the requisite technical expertise, especially given that the development of the theory will tend to be external to the profession.

The ability to derive practical insight within a given professional domain from some external and abstract scientific theory is an important, if intangible, element of professional skill that is particularly important to the actuarial profession given how much of its technical knowledge is developed outside of the profession. But the actuarial profession is far from unique in this regard, and a similar requirement arises in many professional contexts. Consider, for example, how a General may make use of game theory in military strategy and planning; or how an oncologist may make use of radiation therapy in cancer treatments. Ultimately, professional success in these scenarios relies on constructive collaboration between the professional and the theorist. This is sometimes easier said than done, not only because of the inevitable communication challenges between different types of experts, but because it can give rise to a contest over the boundaries of professional domains. This has arguably been a feature of the way actuaries have engaged with economists in recent decades.

A deep and thorough understanding of economic theory can offer a great deal of conceptual insight into many vital topics in actuarial science. Actuaries must still use their professional expertise to make interpretative use of that insight, recognising that the theory is an abstraction, and applying their understanding of the empirical complexities that arise within their professional domain. This

[367] Huber (1996), p. 288.

brings us to the next topic in our philosophical discussion of contemporary actuarial methodology: how professional judgement and skill is and should be treated within actuarial science.

5.7 Professional judgement and modelling skill

In opening the discussion of Pemberton's 1999 actuarial methodology paper, Cartwright closed with the observation that '[philosophy of science] has virtually no accounts available of how to justify concrete causal claims where they do not follow from abstract theory'[368].

Chapter 4.2 noted the significant philosophical problems that can arise in the social sciences when attempting to justify concrete claims even where well-developed abstract theory *does* exist. So, it is perhaps no surprise that philosophy has even greater epistemic difficulties with causal claims in the *absence* of such theory. As per the empirical and applied characterisation of actuarial science developed in Chapter 5.2, we may reasonably reject the very idea that attempts to justify causal claims are a central focus of actuarial science. Nonetheless, Cartwright is highlighting an important point for actuarial science: there is a significant epistemic challenge around how to justify, evaluate or validate elements of actuarial science that she herself has identified as part of its defining characteristics - in particular, the characterisation of actuarial science as skills-based.

Of course, actuarial science is not the only discipline where professional skill or expert judgement is an important and characterising feature of method. Medicine and other types of learned profession may provide other examples. And the extent to which expert judgment features in any professional task may vary by degree. The importance of expert judgement and its role in elevating a task into the domain of a profession is ultimately a function of its cost of failure. A car mechanic can happily 'guess' at the car fault and keep trying until he or she gets it right; a General fighting a war or a neurosurgeon operating on a patient may not get more than one chance to get it right. In such cases, the expert judgment required to reliably make the correct assessment and decision is clearly valuable.

Useful parallels with the role that skill and judgement play in actuarial science may potentially be found in other fields where the use of probabilities play an explicit role in the analytical processes of the discipline. A body of academic research has developed in recent decades which considers how expert judgment may be used in the formation of probability-based risk analysis. The World Bank published a useful review[369] of this literature in 2004, which reported:

"Because of the complex, subjective nature of expert opinion, there has been no formally established methodology for treating expert judgement."[370]

In the absence of such a methodology for handling expert judgment, a popular strategy has been to attempt to aggregate or find consensus amongst many, independently minded experts. This is an intuitive and practical approach to managing the expert judgment problem: perhaps confidence in the use of expert judgement can only be obtained from the judgments of other, independent, experts. (And it may be noted that this outlook is consistent with Chapter 5.5's suggestion that the best way to regulate the use of expert opinion in insurance firms' internal capital models is to publish them, thus making these judgements transparent and available for scrutiny by other experts.)

---

[368] Cartwright in Pemberton (1999), p. 179.
[369] Ouchi (2004)
[370] Ouchi (2004), p. 6.

The World Bank review noted that a range of mathematical techniques have been developed to distil several experts' views on future probabilistic scenarios into a single probability or probability distribution. Beyond these mathematical algorithms, and indeed chronologically before them, 'behavioural' methods such as the Delphi technique were developed. These seek to find a consensus through iterations of anonymous circulation of views amongst a group of experts[371]. The essence of all of these techniques is that it is difficult for a lay person to second-guess an expert, and so it is left to other experts to effectively regulate the opinions and judgements of experts.

The assurances produced by such approaches may not offer a panacea, however. Most obviously, these types of peer group-driven approaches are susceptible to 'group think'. Returning yet again to our 1990s real interest rate forecasting example, it would likely have been quite hard to find many expert opinions that attached a meaningful probability to the actual 2020 outcome for the Sterling long-term real interest rate. The success of these approaches relies on accessing a range of independent expert opinions that span the full spectrum of reasoned expert views, whilst managing to exclude the use of views that are clear 'quackery'. The decision about what does and does not constitute 'admissible' expert judgement is itself an expert judgement. From an epistemic perspective, we find ourselves in an infinite regress familiar in the analysis of inductive reasoning[372].

As noted above, actuaries are not unique in their need to use skill and expert judgment in their professional work. When a physician decides upon a course of treatment, they may undertake a range of test procedures and then consult published charts, formulas and guidelines to translate those results into a 'standard' course of treatment. But the doctor will often use their expert judgement and professional skill to tailor their treatment strategy to reflect the unique circumstances of their specific patient. Indeed, the application of this skill and judgment to unique cases is arguably the essence of what a profession does. For the doctor or surgeon who treats scores of patients, it may be possible in at least some cases for their professional performance to be assessed and monitored. This is arguably harder in the professional domain of actuaries where the success or otherwise of the professional service may take decades to emerge and is subject to many other influences.

The asymmetry of expertise between the professional and their client inevitably creates a need for the client to place some form of faith or trust in their professional advisor, which in turn creates a requirement for the professional to respect that trust by adopting a fiduciary responsibility to act in the client's interests. Professional standards and legal requirements are likely to determine what range of actions may be considered within the boundaries of reasonable profession judgement. But the meeting of such minimum standards does not remove the rationale for the more systematic peer group scrutiny that can be generated by the transparency provided through public disclosure of professional decisions and methods.

Professional skills, complexity, technology and professional jurisdictions

In Susskind & Susskind's influential 2015 book, *The Future of the Professions*, they recognised that expert judgement was a significant feature of the work of any profession. They referred to this professional skill as '*practical expertise*', which they defined as 'the complex combination of formal knowledge, know-how, expertise, experience and skills'[373]. One of the central arguments of Susskind & Susskind is that an internet-driven world will increasingly find ways of centralising and commoditising this practical expertise. Practical expertise, Susskind & Susskind argued, 'traditionally

---

[371] Dalkey & Helmer (1963)
[372] See Chapter 1.3.
[373] Susskind & Susskind (2017), p. 41.

has been held in people's (professionals') heads, textbooks and filing cabinets. Increasingly, this expertise is being stored and represented, in digital form, in a variety of machines, systems, and tools.'[374]

Their book also accepts, however, that this process has a limit: some portion of practical expertise is 'tacit' and not amenable to extraction and digitisation. Such expertise can be informal and heuristic, meaning that there is inherent difficultly in 'mining the jewels from professional's heads'. Nonetheless, Susskind & Susskind write of a 'post-professional world' where the internet-driven world can de-compose, standardise, automate and digitise practical expertise. They argue that automation and Artificial Intelligence (AI) systems can improve the quality of large swathes of professional work. For example, they note a study where algorithmic medical diagnostic systems materially reduced the frequency of false negative diagnoses for breast cancer[375].

Their argument implies that much of professional judgement can ultimately be made redundant. They emphasise, however, that theirs is a multi-decade perspective. Their forecast may or may not ultimately prove accurate. The pace and extent to which professional activities can move entirely from the bespoke crafting of case-specific solutions to fully-routinised algorithms is difficult to predict. But in the meantime, we still have the challenge of how to approach the evaluation of professional skill.

This perspective of Susskind & Susskind is broadly consistent with the (earlier) arguments of Andrew Abbott, the leading sociologist of the professions who is mentioned in the Introduction above[376]. Abbott argued that the complexity of the inferences made by a given profession has a significant impact on the vulnerability of that profession's jurisdiction. Tasks requiring only simple inference will be subject to routinisation and de-professionalisation. The internet-driven world that is the focus of Susskind & Susskind may accelerate this commoditisation of expert skill. Abbott also argued that complexity was not necessarily the best form of defence against this routinisation of professional tasks. Professions whose inferences are always complex and obscure may have their own difficulties. Too much reliance on expert judgement to tackle a series of seemingly idiosyncratic tasks might suggest that the profession's 'abstract knowledge system' (i.e. what we refer to as actuarial science in the context of the actuarial profession) is lacking in scientific legitimacy.

Historically, it has not been unusual for technological developments and the near-inevitable increase in complexity that accompanies them to result in new professions entering the jurisdictions of traditional, established professions (either by usurping them, working alongside them, or acting in subordinate roles to them). For example, in the 19th century, there was virtually no structural engineering profession. Architects did their own engineering. Over the 20th century, the increasing complexity of the technology of building construction led to the emergence of the specialist profession of structural engineers. Today, the architect typically provides the lead direction and retains the overall professional responsibility for a major building project, but essentially delegates some complex professional building tasks to the structural engineers[377]. Clearly, (unpredictable) technology developments can shape the evolution of professions (their size and their activities) in a variety of ways.

Irrespective of how technology impacts on the complexity and commoditisation of the professional tasks of the actuary over the long-term, it seems inevitable that expert judgment and modelling skill

---

374 Susskind & Susskind (2017), p. 109.
375 Susskind & Susskind (2017), p. 48.
376 Abbott (1988)
377 This example is based on Abbott (1988), p. 73.

will play a meaningful role in actuarial science over the foreseeable future. The philosophical difficulty identified by Cartwright at the start of this section probably does not admit of any perfect solution. But actuaries must perform in an imperfect world. The key to the necessary oversight of actuarial expert judgment is transparency and disclosure. This applies to both the communication of the methods of actuarial science to others outside the profession who have expertise in the techniques employed therein; and to the specific expert judgments that actuaries make in the course of their work, which should be made available to other actuaries and other relevant experts through appropriate disclosures wherever possible.

6 Implications for an improved actuarial methodology

Chapter 5 developed an analysis of some features of the contemporary methodology of actuarial science. The discussion identified some of the key characteristics of actuarial science, as determined by the nature of the questions it sought to answer and the nature of the socio-economic phenomena that it analysed in doing so. One of the central points to emerge from this analysis was that there is a substantial epistemic limit to what can be predicted (deterministically or stochastically) for the future behaviour of most of the phenomena of actuarial interest. When modelling these phenomena, this inevitably leads to what is commonly known as model uncertainty. The level of uncertainty that applies to the long-term nature of many phenomena of actuarial interest means that much of actuarial modelling is conducted under conditions of deep model uncertainty.

The presence of deep model uncertainty does not imply actuarial models are useless or that actuaries have made significant errors in their model-building. Rather, it was argued in Chapter 5 that this unavoidable deep model uncertainty has fundamental implications for *how models are used* and *what questions are capable of being reliably answered*. In the presence of deep model uncertainty, actuarial models are best suited as insight-generating tools for the use of actuarial experts who understand them, rather than as answer-generating machines that are built by actuarial experts to generate answers for lay clients. Put another, in the language of the philosophy of social science as discussed in Chapters 3 and 4, in the presence of deep model uncertainty, actuarial models are best employed as part of an interpretivist rather than positivist methodology.

This chapter considers some of the implications of this perspective for the methodology of actuarial science. We will consider some specific topics in contemporary actuarial science and discuss how the above perspective implies some alternative approaches to those commonly used by actuaries today. These topics are chiefly concerned with some form of measurement: of risk, or capital or some form of funding requirement. Before considering these topics in more detail in Chapter 6.2, Chapter 6.1 first considers another question where the above perspective on the presence of deep model uncertainty may have important implications: what forms of risk naturally reside on the balance sheets of actuarial financial institutions and what form of risks are best accumulated elsewhere (by individuals or more directly in the capital markets, for example). Finally, as part of this theme of looking to the future and considering how the methodology of actuarial science can develop, Chapter 6.3 discusses how the emerging technological capabilities of big data and data science can be considered within the methodological framework developed here.

6.1 Risk appetite in actuarial financial institutions

A discussion of risk management might naturally begin with some sort of specification of the range of risks that are to be retained and thus must be managed. This is often referred to as specifying the risk appetite of the individual or institution in question. Risk appetite is usually expressed quantitatively. That is, *how much* of a given type of risk is an institution or individual willing to bear. But here we will begin with the more basic, binary question: *which* risks should an actuarial financial institution be willing to bear; and, equally importantly, which risks naturally best belong somewhere else.

These questions might be considered beyond the domain of actuarial science. It might be argued that it is the actuary's role to advise their client on how to measure and manage whatever forms of risk the client wishes to be exposed to. Irrespective of where we might choose to define the boundaries of actuarial science, here we will consider what the above methodological analysis says about the types of risk that a financial institution such as an insurance firm or defined benefit pension fund ought to seek to be exposed to, and which types of risks they ought to avoid. For brevity, in this discussion we will refer to these types of institutions as actuarial financial institutions; and we will use the term 'balance sheet' in the broadest sense to refer to the assets and liabilities of the institution, irrespective of the particular legal structures that pertain.

We begin with a general assertion that carries significant implications for risk appetite: *risks that are subject to deep uncertainty are not risks that can be efficiently managed or sustainably accumulated by actuarial financial institutions such as insurance firms or pension funds. Actuarial financial institutions should focus on facilitating the pooling of diversifiable risks and should be strongly averse to the accumulation of non-diversifiable uncertainties*.

This might seem a bold statement, particularly in the context of Chapter 5, which concluded that much of what actuaries measure and manage (such as longevity risk and economic and financial market risk) is subject to deep uncertainty. The logic of this assertion will be developed using examples developed throughout Chapter 6.1. But such an assertion gives rise to a couple of immediate obvious questions…if actuarial institutions avoid those uncertainties, what is left for them to do? And haven't actuarial financial institutions such as insurance companies borne these uncertainties very successfully for hundreds of years, generating an enviable long-term track record in consistently delivering on their promises in the process?

A Historical Perspective on the Long-term Resilience of Insurance Firms

History suggests a more nuanced reality than that expressed by the above claim. Consider life assurance in the UK, for example. It is certainly true that UK life offices have accrued an enviable long-term record of making good on policyholder promises. How has it done so? For two centuries of the roughly two-and-a-half centuries of actuarial history, the mutual life office has been the dominant form of actuarial financial institution in UK life assurance. The financial structure of the British mutual life office, as developed and implemented at the end of the eighteenth century by pioneers such as Richard Price and William Morgan[378], was carefully and deliberately designed to eschew exposure to long-term uncertainties on its balance sheet. This is what made the mutual life office a resilient, sustainable and successful financial structure for two hundred years.

Specifically, the with-profit policy, as configured then, facilitated the pooling of individual mortality risk, whilst passing back the bulk of the exposures to the long-term uncertainties in interest rates, investment returns and longevity trends back to the with-profit policyholders. Non-profit policyholders did manage to pass all of these risks to the institution (which was wholly owned by the with-profit policyholders), but the offices were careful to ensure that the bulk of policyholder liabilities were of the with-profit form, thus ensuring that the exposures arising from the non-profit book could be comfortably absorbed by the risk-taking nature of the with-profit policies[379]. Moreover, for the first 200 years of the history of the mutual British life office, the assets backing non-profit liabilities were invested in very low risk assets (mainly government bonds). Allocations to risky asset classes such as equities were limited to excess assets that were not required to match liability guarantees (arising from either with-profit or non-profit liabilities).

This structure worked very well for almost two hundred years. It was only in the second half of the twentieth century that the British mutual life offices started to develop major exposures to interest

---

[378] See, for example, Turnbull (2017), Chapter 2; and Ogborn (1962), Chapters VII, VIII and IX for further discussion of the historical development of the mutual British life office.

[379] In his essay on the consequences of uncertainty for actuarial practice, *Prescience and Nescience*, Frank Redington seems to go a step further and argues that our inability to forecast interest rates and mortality rates meant that long-term non-profit policies were 'unreasonable' and asked 'should we not abandon them except in very special circumstances'. (p. 533).

rate risk (in the form of guaranteed annuity option liabilities, for example) and to financial market risk (by investing in equities and real estate beyond the level that permitted guarantees to be fully matched by low risk bonds). Whilst the flexibility in the pay-outs to with-profit policyholders could absorb some of the impact of these risk exposures, some of the impact was inevitably left on the balance sheets of the mutual life offices. And the eventual law courts' judgement on what constituted the reasonable expectations of with-profit policyholders clarified that the flexibility available in with-profit pay-outs was less than actuaries and the management of mutual life offices had previously understood[380].

So, the historical long-term resilience of the mutual British life office was, at least in large part, due to limiting the exposure to long-term uncertainties such that they could be structurally absorbed by its with-profit policyholders via the explicit participation of their policy pay-out in these risks. Meanwhile, this structure successfully transferred and pooled the diversifiable element of all policyholders' mortality risk (with-profit and non-profit).

What of today's life assurance firms? Well, by contrast to the first 200 years of the history of British life offices, most of today's firms bear a lot of exposure to long-term uncertainties. This comes most obviously in the form of their investment strategies, where the assets backing fixed non-profit liabilities tend to be invested in risky asset strategies (and the current trend appears to be towards assets with more risk, more complexity and less liquidity). Today's insurance firms also bear significant amounts of long-term uncertainty in the form of exposure to long-term longevity trends (which mainly arise in the form of (non-profit) fixed annuity policies).

Below we further discuss the rationale (or lack of it) for insurance firms bearing these two forms of uncertainty.

Insurance Firms and Longevity Protection

Many life assurance firms strategically seek exposure to long-term longevity uncertainty. This exposure has historically been obtained in the form of non-profit life annuities. Since the removal of compulsory pension annuitisation in the UK in 2015, the demand for such policies has fallen substantially. Today, another significant source of long-term longevity exposure for life assurers arises from Defined Benefit (DB) pension funds who are seeking to transfer their exposure to this form of risk to an insurance firm (DB pension fund topics are discussed further below).

Individuals with Defined Contribution (DC) pension funds are currently choosing to bear significantly more individual longevity risk in their retirement than has been the case in recent past generations. It seems counter-intuitive for individuals to bear a risk that is highly material to them and that, at least in large part, can be easily pooled and diversified. Why is the conventional solution of an annuity contract not working here (in the sense that it is not attracting significant demand from DC pensioners)? The reasons for this may be numerous and complex and some of them may go beyond only actuarial considerations. In the context of the above discussion, however, it is natural to ask: is the deep uncertainty associated with long-term longevity trends a relevant factor?

When considering the longevity outcome for an individual, it can be de-composed into two parts:

- There is the overall longevity experience of the population (or relevant homogenous sub-group of the population that the individual can be placed in). As was anecdotally demonstrated and discussed in Chapter 5, the long-term prediction of this experience is subject to deep uncertainty. When an individual retires at, say, age 65, it may well turn out

[380] See, for example, Turnbull (2017), Chapter 5; and Boyle and Hardy (2003) for further discussion.

that the actuarial life expectancy for the (relevant sub-group of the) population at that point in time turns out to have been wrong by a few years. We do not have a reliably accurate way of measuring just how wrong these estimates could be, but even with a sceptical outlook on positivist social science methods, we can have some confidence that this estimate will not be wrong by, say, 20 or 30 years for the population of 65 year olds.

- Then there is the longevity outcome of the individual *relative* to the average experience. This part of the individual longevity outcome *could* easily differ from the average life expectancy by 20 or 30 years. This part of longevity risk is statistically independent of the experience of the relevant homogenous population group. The outcome of accumulating this form of risk is therefore *not* subject to deep uncertainty. It is a diversifiable risk.

The argument above implies that life assurers should not write policies that accumulate the first component of individuals' longevity risk, as it is subject to deep uncertainty. If the insurer does accumulate exposure to this uncertainty, they will have to hold significant capital to support this exposure. The holding of this extra capital is costly and the associated cost will naturally have to be passed onto the policyholder in the form of an increase in the policy premium. The bearing of this risk will also require some educated guesswork in the setting the longevity assumptions in the policy pricing, and the insurer may err on the side of caution when setting these assumptions.

From the individual's perspective, it makes little sense to pay for the costs associated with passing the long-term uncertainty of average longevity trends to the life assurer: for the individual, the vast bulk of their risk can be represented by how their longevity experience *differs* relative to the average. This can be cheaply diversified by an insurer. It is the other part, the average trend risk element, that incurs the greater costs.

This analysis suggests that the traditional non-profit annuity is not a particularly good longevity risk solution for retirees. By transferring both the idiosyncratic and long-term average trend components of their longevity risk to the life assurer, the annuity provides a longevity protection solution that is more expensive than is required to meet the key individual need. A sustainable and cost-efficient long-term longevity solution would be one that delivers a form of risk transfer in which the insurance firm provides the policyholder with protection from idiosyncratic mortality risk (the diversifiable bit; that is, the risk that a newly-retired individual lives to 105 when the average experience is to live to 88), whilst returning to the policyholder the long-term longevity trend risk (the non-diversifiable, uncertain bit; that is, the risk that the relevant population lives to an average age of 91 when it was expected at the time of inception of the risk transfer that the average outcome would be living to 88).

There are many forms of insurance product design that could potentially achieve these two objectives of facilitating the pooling of the diversifiable part of longevity risk, whilst returning the non-diversifiable part to the policyholder. Such solutions would remove the bulk of the policyholder's longevity risk without requiring the insurance firm to bear the long-term uncertainties that inevitably entail very subjective pricing and holding substantial and costly capital to support these risks.

Insurance Firms and Investment Risk

Why do insurance firms often have a significant financial market risk appetite for the assets backing their insurance policyholder liabilities[381]? The answer seems very straightforward: because that risk is well-rewarded, and those investment rewards are often a major part of insurance firms' revenues and profits. But according to basic economic logic, no obvious shareholder interest is served by investing these portfolios in risky assets rather than risk-free assets – in finance theory, shareholder wealth is not enhanced by switching balance sheet assets from government bonds to corporate bond, equities or anything else. Economic theory would suggest that, in the absence of complicating factors such as tax, insurance investment strategy choice is irrelevant to insurance company shareholder value[382]. The basic conceptual point is that the insurance firm's shareholders are quite capable of investing in these risky assets directly without incurring the extra costs of doing so via an insurance company. There is therefore no economic rationale for the insurance firm to have a significant financial market risk appetite when investing such assets.

This simple argument runs quite contrary to prevailing orthodoxy. It begs the question: if the purpose of insurance firms is not related to taking investment risk on behalf of shareholders, what then *is* their purpose? As argued in the discussion of longevity risk above, insurers efficiently facilitate the pooling and diversification of policyholders' diversifiable risks. This risk-pooling is a very useful economic activity. It is also a relatively straightforward function that we would expect to be associated with low risk to insurance shareholders and low profit margins. Of course, most insurance firms today are more than 'just' insurance firms. They provide a broader array of financial services to their customers than the pooling of diversifiable risk. But that does not alter the economic logic that argues that no obvious shareholder interest is served by taking financial market risk with the assets backing insurance liabilities. Of course, beyond this basic and general economic argument, there are some potentially important complicating factors that arise in the context of an insurance firm. These factors ought to be considered carefully in this discussion: they include *liquidity*, *regulation and leverage*[383]. Let's briefly consider the possible effects of each of these on insurance asset strategy.

When a policyholder purchases an insurance policy they are, sometimes, buying something that is very illiquid, in the sense that the policy does not have a readily realisable cash value. This is not always the case – many forms of insurance policy provide the policyholder with the option to surrender or sell their policy back to the insurance firm prior to its maturity. But there are some forms of insurance policy – life annuities being the obvious example – where the policyholder has no such surrender option. In this case, the policyholder has purchased a highly illiquid asset. If we accept the premise that illiquid assets are valued at less than otherwise-equivalent liquid assets, this has two immediate implications: insurers should take advantage of this illiquidity on the liability side of their balance sheet by investing in illiquid rather than liquid assets[384]; and policyholders should

---

[381] Here, we are referring to policyholder liabilities whose cashflows are not linked to the investment performance of the assets.

[382] This is not a new idea, either in corporate finance theory or in the specific context of insurance investment. For an example of discussion of the insurance investment irrelevance proposition in the actuarial literature see Section 6.3 of Bride and Lomax (1994). This logic has also been applied to the analysis of the investment strategy of Defined Benefit pension funds (see Exley, Mehta and Smith (1997)). DB pension investment strategy is discussed further below.

[383] There are also other important factors such as the tax treatment of different assets and how the tax treatment that applies to investments held by insurers differs from that which applies to other forms of investor. Such factors are doubtless important, but their economic implications are more straightforward, and in in the interests of brevity the topic of tax is not discussed here.

[384] This is another not-new idea. It dates back in the actuarial literature at least as far back as Bailey (1862) (no typo!). See Turnbull (2017) Chapter 3 for a fuller discussion.

require a discount in the pricing of the insurance policy to compensate them for its illiquidity. This provides a rationale for insurers having some appetite for asset illiquidity (and passing any available illiquidity premium on to the policyholders in the form of a reduced insurance premium). This asset illiquidity appetite is distinct from an appetite for financial market risk, though it might be argued it is difficult to obtain material illiquidity premia (after costs) without being exposed to some market risk.

Insurance firms in most major jurisdictions today operate under a regulatory system that includes a risk-sensitive solvency capital requirement. This means that when insurers' asset investment risk is increased, shareholders will be compelled by regulators to hold more capital on the insurance balance sheet. It is generally accepted that the holding of this capital on the insurance balance sheet incurs a cost for shareholders. This cost is related to the frictional costs incurred by shareholders in tying up their capital on an insurance balance sheet (costs such as double taxation, management agency costs and the costs of financial distress)[385]. This suggests that, all other things being equal, the investment irrelevance proposition noted above should be taken a step further: it is not merely the case that shareholders should be *indifferent* to asset risk-taking on the insurance balance sheet, they should have a preference for *less* asset risk-taking on the insurance balance sheet (as this risk-taking creates a cost of capital that could be avoided if the shareholder instead obtained these risk exposures directly).

Nor does the taking of investment risk appear to be obviously in policyholders' interests. Someone else is taking risks with their insurance premiums. The policyholder does not participate in the investment upside[386]. But the downside risk makes their promised benefits less secure (additional capital requirements notwithstanding). This suggests that the investment risk-taking could be a lose-lose for shareholders and policyholders. But almost all insurers today have a material appetite for investment risk. Why?

Economic theory can offer a form of explanation once it is recognised that the insurance balance sheet is *leveraged by borrowing from policyholders*. The shareholders' equity claim on the firm can therefore be viewed as a call option on the firm's assets – the shareholder gets whatever asset value is left over after creditors (including policyholders) have been paid, but limited liability means this amount cannot be negative. Viewing equity as a call option on the firm's assets is certainly not a new idea. It dates back at least as far as the Modigliani-Miller theory of the 1950s and it was a major motivation for the development of option pricing theory in the early 1970s.

Considering the policyholders of an insurance company as lenders to the insurance firm is also not a new idea[387]. Policyholders own a debt of the insurance firm (in the form of an insurance policy that obliges the insurer to pay the policyholder when specified events or circumstances arise). As a debtholder, they are exposed to the risk that the insurance firm's assets will prove insufficient to meet their claim if and when it falls due. The policyholder is short a put option on the firm's assets.

From this perspective, an increase in asset risk (volatility) represents a transfer of wealth from policyholder to shareholder (all other things remaining equal): it increases the value of the shareholder's call option on the firm's assets; and there is an equal and opposite reduction in

[385] See, for example CRO Forum (2006).
[386] Clearly there are forms of insurance policy where the policyholder does participate in the investment upside, but here we are discussing the asset strategy for assets backing liabilities with no such linkage
[387] Again, see Bride and Lomax (1994), for example.

policyholder value that arises from the increase in the value of the put option that the policyholder is short.

This reduction in the value of the insurance policy could be incorporated into the pricing of the policy in the form of a reduction in the premium charged up-front to the policyholder. Such a reduction in the insurance premium would ensure that both the shareholder and policyholder share in the (potential) economic rewards that are associated with the firm's chosen amount of investment risk-taking. The shareholder obtains a higher expected, but riskier, reward, and the policyholder still receives an insurance pay-out that is fixed with respect to investment risk, but which is now larger than it otherwise would be per £1 of premium (but which now comes with a greater risk of default). Insofar as the shareholder can increase risk without reducing the policyholder premium payable for a given insurance policy, the shareholder has an incentive to increase risk.

Some may object to the rather invidious-sounding characterisation of insurance firms' investment risk-taking as an attempted transfer of wealth from policyholder to shareholder. It could be argued that the policyholder benefits from this risk-taking as well as the shareholder – the shareholder shares the prospective proceeds of the risk-taking with policyholders in the form of an up-front (fixed) reduction in the premium for the insurance. So, is investment risk-taking actually a win-win? The central point of the above analysis is that it is only right that the policyholder is charged less for fixed promises when the shareholder takes investment risk with the policyholder's premium. After all, their policy is worth less in the presence of the default risk created by the investment risk (albeit the policyholder may not fully appreciate that fact). Whether policyholders want a cheaper insurance policy that comes with more default risk is an interesting and open question that must ultimately hinge, at least in part, on how such products are sold and regulated.

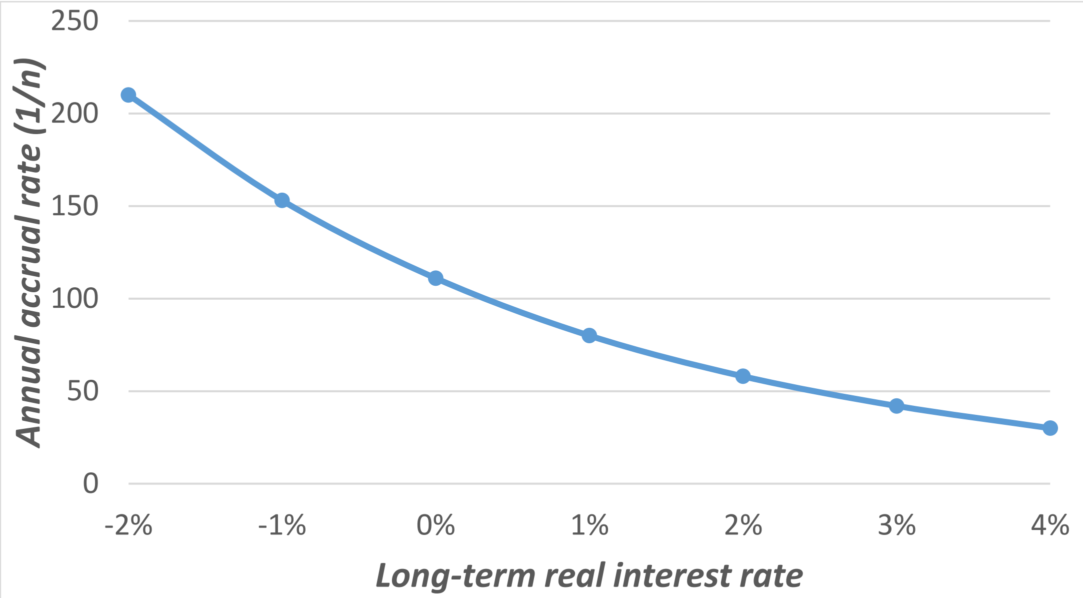
It might also be argued that insurance firms' investment risk-taking with assets backing insurance liabilities is important for the wider economy. These funds are used to finance critical infrastructure and such like. If all of those assets were invested in risk-free bonds, the argument goes, they will simply make risk-free bonds even more expensive than their current all-time record levels. It could be argued in response that if insurance firms wish to participate in risk-taking in the real economy, they should design insurance products that produce a balance sheet liability structure that is fit for that purpose. This likely means allowing the policyholder to directly and explicitly participate in the investment returns of the assets (such as in unit-linked or with-profit-style product structures). Meanwhile, de-risking the investment strategy for 'non-profit' insurance liabilities would release risk capital, reduce investment management expenses and reduce the material costs of regulatory compliance that are associated with holding those assets under principle-based solvency systems such as Solvency II.

The Defined Benefit Pension Fund Concept

Specific approaches to assessing the funding adequacy of Defined Benefit (DB) pension funds, and their implications for investment strategy, will be discussed in Chapter 6.2. In the context of the current discussion of the risk appetite of actuarial financial institutions, however, it may be noted that the cost of future accruals of a DB pension depends on the future levels of phenomena such as interest rates, salary inflation, price inflation and longevity changes, all of whose behaviour is deeply uncertain. It may be possible to put in place investment strategies that can largely or partly mitigate the future variability of the costs associated with past accrued benefits, but this is much harder to do for future accruals. Variability in the future contribution rate required to meet the economic cost of newly accruing defined benefits is largely unavoidable, and the quantum of that long-term variability is subject to deep uncertainty.

We can illustrate the sensitivity of the cost of DB pension benefits to movements in deeply uncertain variables such as the real interest rate by means of a simple example. Consider, for a scheme at inception, the rate of pension accrual that implies an economic value for the pension liability that is equal to the value of the stream of annual contribution rates set at 15% of salary. Exhibit 6.1 shows how this varies as a function of the real interest rate[388].

Exhibit 6.1: Break-even accrual rates as a function of the long-term real interest rate with a contribution rate of 15% of salary



The chart highlights how conventional DB pension accrual rates are unlikely to be economically sustainable in low real interest rate environment. DB pension funds were simply not designed to work with negative real interest rates, and the chart suggests that an accrual rate of less than 1/100th would be necessary to make them sustainable with a 15% salary contribution in a negative real interest rate environment. There is no natural mechanism, other than increasing contributions, available within the DB pension framework to respond to the increased costs that arise in such circumstances. In this illustrative example, an annual contribution rate of 58% of salary would be required to finance a 1/40ths accrual rate in the presence of a real interest rate of -1%.

As we saw in Chapter 5, until quite recently actuaries, and indeed probably almost everyone else, viewed sustained long-term negative real interest rates as virtually or indeed literally impossible. If the long-term real interest rate behaved in the way described in Wilkie's 1995 model (see Exhibit 5.2 above) - that is, it always remained within the range of 3%-5% - then DB pension finance with a 1/40ths accrual rate is sustainable and really quite straightforward. But, as discussed in Chapter 5, there was no epistemic basis for confidently assuming real interest rates will behave in this way indefinitely into a multi-decade future. Once we accept the reality that real interest rates of -2% are as possible as real interest rates of +2%, the scale of the potential variability in the costs that the sponsor is required to bear in providing its employees with a DB pension is enormous in the context of overall remuneration and staffing costs.

---

[388] This simple illustration considers a male joining the pension fund at age 25 and retiring at 65. The mortality rate basis assumes a(90) without mortality improvements. Real salary growth is assumed to be zero and the pension is assumed to be inflation-linked.

Exhibit 6.1 highlights a broader, and even more fundamental challenge to pension finance. Irrespective of the specific form of advance funding of pension provision – Defined Benefit, Defined Contribution or something else in the middle – the economics of saving for a pension are profoundly challenged by a negative long-term real interest rate. If such economic conditions are to prevail for a long period of time, there seems no obvious alternative financial solution to the affordable funding of pension provision than for individuals to work much later into life than has been the case in recent decades.

To be clear though, it is not being argued here that negative real interest rates are in some sense inevitable for many years to come. The arguments of Chapter 5 highlight the folly of such an assertion. Perhaps the only basis for optimism is to note that uncertainty works both ways – whilst it may not seem that way given the history of interest rate movements over the last forty years, future real interest rate changes may surprise on the upside as well as the downside. Whatever the proposed financial solutions to our current pension challenges, the actuarial analysis should not assume that deep uncertainty in the outlook for real interest rates and other important economic and demographic variables does not exist.

6.2 Risk and capital measurement

Chapter 5 discussed some of the ways in which actuarial models are used in the development and delivery of the professional output of actuaries. The chapter also argued that there were at least some areas of current professional practice (such as principle-based capital assessment) where alternative governance structures and transparency standards could lead to improved outcomes. Central to this argument was the idea that the deep uncertainty associated with the socio-economic phenomena of actuarial interest made models more useful as interpretative tools in the hands of experts who deeply understand them, rather than as answer-generating machines developed by experts to deliver answers to lay clients or authorities.

The discussion recognised that this created a problem of how to oversee experts and expert advice. Indeed, this was the one of the issues that resulted in the use of actuarial models changing in the directions that it has done in recent decades. We argued that, in the topical example of principle-based regulatory capital assessment in insurance, extensive peer review through public disclosure of detailed actuarial reasoning would be a better form of quality assurance than the private internal model approval process used in systems such as Solvency II today. We also noted this was generally consistent with the broader academic literature on oversight of expert advice in quantitative modelling, which is focused on obtaining a form of consensus amongst experts as the primary means of overseeing the quality of expert advice.

This section considers some more specific topics in quantitative actuarial reasoning and develops some methodological suggestions that may have contemporary actuarial relevance. Before doing so, we may note that in the financial world advocated in Chapter 6.1, actuarial risk measurement would be more straightforward than it is today - as many of the exposures to deep uncertainty that are accumulated by today's actuarial financial institutions would not be present on their balance sheets in a Chapter 6.1 world, at least not to the degree they are today. Instead, the vast bulk of risks on an actuarial financial institutions' balance sheet would be diversifiable and would indeed be diversified away on the balance sheet. The institution's function would be to facilitate efficient, orderly and equitable pooling of diversifiable risks. The challenge of attempting to measure the consequences of accumulated exposures to deep uncertainty would therefore be largely avoided in this setting.

The two topics below, however, reflect the world we are in today, and the world we will continue to have to manage given the legacies of long-term business created in the past. And this means wrestling with the impossible task of measuring deep uncertainty. Solving the impossible is, thankfully, beyond the scope of this work. Nonetheless, the following analysis will use some of the insights garnered in this work's philosophical reflections to attempt to identify satisfactory methodological approaches in the presence of deep uncertainty.

The measurement strategies that are developed in both the case studies below have one key aspect in common that may be particularly apparent if you recall the discussion of the anatomy of actuarial models in Chapter 5.2. There, it was noted that actuarial models can be usefully considered as having two elements: the first being concerned with the projection of future cashflows, and the second with the transformation of those cashflow projections into some form of present value metric through the use of a form of discount function. The approaches developed in both the examples below can be viewed as approaches to finding a form of discount function that makes the present value metric *more robust and less sensitive (though still not completely insensitive) to measurements of deep uncertainty*. Put another way, in both cases we try to answer questions that *are more capable of being reliably answered* than the conventional questions of actuarial consideration.

Case Study (1) – Defining the quantitative metric for risk-based capital assessment in insurance

Actuaries in the UK have been assessing probabilistic risk-based capital requirements for at least some types of long-term guarantees found in life business since around 1980[389]. Different approaches have been taken to defining a risk-sensitive probabilistic capital measure, and these different definitions imply a different set of calculations are required to assess the defined measure. These different approaches are discussed further below. Before doing so, we will first note they each have some key features in common.

The probabilistic capital requirement is assessed by estimating the capital required to be held today such that the probability of some adverse future solvency outcome occurring at some future point(s) in time (which could be a specified interval of time; any time in the future; or after the final liability cashflow is paid) is no greater than some specified level, given the risks on the balance sheet and how they are managed. So, the approaches differ in their definitions of what constitutes an adverse solvency event, over what time horizon it is considered, and what level of probability the capital requirement should support for the adverse solvency event. In all cases, this assessment is made using some form of stochastic model, and this is usually implemented using Monte-Carlo simulation. Irrespective of the time horizon of the capital measure, the models may make stochastic projections of asset returns and other relevant phenomena over the lifetime of the liabilities (for liabilities whose value depends on future asset returns). These stochastic projections are used to infer some form of probability distributions for the liability cashflows that may arise at different times. These probability distributions are then transformed into some form of present value metric as required by the capital definition.

There are different ways of specifying the present value discount function used in the above process, and these correspond to different definitions of the adverse solvency outcome that is used in the risk-based capital assessment. The origins of the 'traditional' (in the UK) actuarial approach to risk-based capital assessment can be traced back to the development of risk theory in Scandinavia in the years following the end of the Second World War[390]. These ideas were applied to the long-term financial risks associated with the long-term maturity guarantees of unit-linked business by Sidney

---

[389] See Turnbull (2017), p. 194-204 for historical background.

[390] See Turnbull (2017), Chapter 7, pp. 286 – 292 for some historical background on the actuarial development of risk theory.

Benjamin, a leading UK actuary, in the 1970s[391]. This approach to defining the capital requirement for a guarantee calculates the amount of assets that should be held today such that, along with the accumulation of any regular premiums charged for the provision of the guarantee, the probability of having insufficient assets to meet all guarantee cashflow shortfalls as they fall due is no greater than a specific size (and this size of probability may be referred to as the probability of ruin). This approach is sometimes referred to as the 'run-off' approach. It was officially sanctioned by the Institute and Faculty of Actuaries in 1980 when a Working Party tasked with considering the topic of reserving for these long-term maturity guarantees advocated this approach (with a probability of ruin of 1%)[392].

The modelling mechanics of the run-off approach requires the probabilistic projection of the returns of the assets that the guarantee is written on, as well as the returns for the investment strategy pursued for the capital and the premiums charged for the guarantees, over the full time horizon of the guarantee. The cashflows shortfalls can then be projected as a function of the underlying asset's investment returns, and discounting using the returns that are associated with the assumed investment strategy for the capital. With a 1% probability of ruin, the 1st percentile of the probability distribution of the present value then defines the capital requirement. This run-off approach, based on assessing the probability of being able to fund all liability cashflows as they fall due given a starting level of assets and investment strategy, naturally extends to any form of liability. Again, the approach will require (joint) probability distributions for the behaviour of all phenomena that may materially impact on the cashflows of the assets and liabilities over the full term of the liability cashflows. *The content of Chapter 5 (Exhibits 5.2 and 5.3, for example) may inculcate a certain sense of concern at the prospect of being professionally required to produce a reliably accurate estimate for such a capital definition.*

An alternative probabilistic definition of a risk-based capital requirement, Value-at-Risk (VaR), emerged in the banking sector in the 1990s, and was notably adopted in the European Union's Solvency II regulatory framework for insurance firms when it was implemented in 2016. It is also used in the International Association of Insurance Supervisors' proposed Insurance Capital Standard[393].

The key idea underlying the VaR approach is that the capital requirement is based on an assessment of the cost of transferring the liabilities to a third-party in a stressed near-term scenario. This is calculated in two stages:

- First, by calculating the cost of transferring the liabilities to a third party today;
- Second, by probabilistically projecting how this cost may change over the specified near-term projection horizon;
- And, finally, by finding the capital required to fund the increase in cost that arises at the end of the projection horizon with some specified level of probability (note the amount of capital required to fund the increase in cost will depend on how the assets backing the liabilities are invested; for example, if a reserve equal in value to the current cost is invested in a perfect hedge, no capital will be required in excess of this reserve as the end-period asset value will be equal to the end-period liability value in all states of the world).

[391] Benjamin (1976)
[392] Ford et al (1980)
[393] IAIS (2018)

The first calculation – calculating the cost of transferring the liabilities to a third party today - uses insights from the developments in option pricing theory of the 1970s and the more general economic valuation theory that followed it in the 1980s to specify a stochastic discount function that, when applied to the stochastic cashflow projection of the liabilities over the full term of the liability cashflows, will deliver the current 'market-consistent' cost of the guarantee. This is often referred to as 'risk-neutral valuation', which is a somewhat unfortunate and confusing misnomer, as the essential idea of the whole approach is based on the theoretical insight that the valuation is the same in a risk-neutral, risk-averse and risk-seeking world. But that is a point of technical detail. The essential point is that the cost of transferring the liabilities to a third party today is estimated with reference to the current market prices of relevant assets. There will usually be an insufficient availability of such assets to definitively determine the 'market-consistent cost' of the liabilities, and some extrapolation, interpolation and estimation will therefore usually be necessary on the part of the actuary performing the valuation.

The second and third parts of the process involves projecting this cost, and the assets backing it, over a short period of time and estimating the assets required such that the market value of the assets exceeds the market-consistent cost of the guarantee with some specified probability level. In Solvency II, the time period is 1 year and the probability level is 99.5%. One rationale for the assumed time period is that it represents the length of time it would take to practically implement the transfer of the liability to a third-party.

In the discussion above, we have supposed that both the run-off and VaR approaches use a specified percentile level as the probabilistic 'tolerance' level. Numerous technical studies have considered whether a percentile point is as statistically robust as other tail measures such as Conditional Tail Expectations (CTE) in defining the probabilistic measure of risk-based capital requirements[394]. This consideration of how to define the tail risk arises under both the run-off and VaR capital definitions. That is, under either approach, a *y*% CTE could be assessed instead of using the x^{th} percentile.

These two definitions of risk-based capital requirement will behave quite differently. They may exhibit distinctly different sensitivities to changes in financial market prices. They also rely on different assumptions. One of the arguments forwarded in favour of Solvency II's adoption of a 1-year VaR definition instead of the run-off approach was that it did not rely on so many subjective parameter choices, and hence the 1-year VaR approach would be more objective and scientific. For example, the run-off approach is highly sensitive to long-term assumptions for asset risk premia and its path behaviour, and the VaR assessment is not sensitive to this assumption to the same degree. Nonetheless, the estimation of a 1-year 99.5^{th} percentile estimate for an equity return is not self-evidently more reliable than a 95^{th} percentile estimate of the 30-year equity return. In both cases, the same basic difficulty arises: the instability and non-stationarity of our human world means the estimation of these types of probabilities are intrinsically unreliable.

Although Solvency II did not come into force until 2016, its conceptual framework was established in the early 2000s. Since the global financial crisis of 2008, there has been a shift in intellectual sentiment (if not actuarial practice) away from the market-consistent VaR approach and back towards something more like the traditional actuarial run-off approach[395]. But we are fooling ourselves if we think either of these definitions of risk-based capital requirement can be statistically measured in any meaningful, reliable sense. The debate over which approach is better is

[394] See, for example, Hardy (2006) for a comprehensive introductory treatment in an actuarial context.
[395] See Turnbull (2018) for a fuller discussion of this change in sentiment that has been voiced by some leading central bankers, policymakers and industry leaders.

interminable because neither of the alternatives under consideration are satisfactory. The consensus ranking of market value and run-off solvency approaches will therefore go through cycles where one approach is tried for a while, then inevitably found to be unreliable, thereby leading to a switch in the consensus view so that the other is then regarded as being quite self-evidently superior, and the cycle will repeat again. Meanwhile, debates about the mathematical and statistical properties of VaR versus CTE when applied as risk measures for financial market risk are like arguing over how best to estimate how many angels can dance on a pinhead. This conclusion leads to the obvious question: but is there anything better?

Before attempting an answer to that question, it is important to clarify a key point in this discussion: the substantial epistemic limitations of both the run-off and VaR approaches to risk-based capital assessment do not imply that these forms of modelling analysis are of no value to the actuary. On the contrary, both the run-off and VaR definitions of risk-based capital requirements can provide useful and insightful models of capital requirements that can be informative in the hands of an expert who has a deep understanding of the models and their limitations.

To illustrate this point, recall the historical episode of the Maturity Guarantees Working Party and the assessment of capital required to back those guarantees using a run-off approach. This work, and some of the actuarial research that preceded it, highlighted a vital truth: that life assurers at the time were writing very significant long-term guarantees and did not appreciate the level of non-diversifiable financial risk that these guarantees generated. The stochastic modelling used by the Working Party robustly demonstrated that this risk was undoubtedly material and that reserving methods ought to be updated to better reflect these risks. However, having obtained this insight, actuaries could have been left to use their own expertise and judgement to determine an appropriate reserving method, rather than having one essentially prescribed. The prescribed method was, naturally, full of limitations, and the prescription reduced actuaries' incentive to further innovate, learn and improve the capital requirement-setting process.

Likewise, a VaR analysis can also bring its own insights into the form and quantum of risk exposure and the amount of capital required. The two capital measures can be especially insightful when considered in conjunction with each other. For example, if the run-off capital model implies a substantially smaller capital requirement than the VaR assessment, this may imply the result is especially dependent on assumptions that do not materially affect the VaR assessment (such as long-term risk premia assumptions or asset return mean-reversion). These alternative capital definitions and implied modelling approaches can be viewed as complementary diagnostic tools for the actuary, rather like, say, how both MRI scans and blood tests may be used by a doctor: they provide diagnostic evidence to the expert, who considers all of the available information at hand to make the diagnosis and prognosis for the unique case under consideration.

Let's now return to the question: is there a probabilistic definition of a risk-based capital requirement that is in some sense better than the VaR and run-off approaches? Here an approach is proposed that is arguably superior to both the VaR and run-off approaches, and superior in two key respects: it has greater economic meaning and consequence than the other two approaches; and it relies less on attempts at probabilistic measurement of the deeply uncertain.

The basic idea of this approach is based on the recognition that, in the context of an insurance firm with non-linked liabilities, risk-taking (for example, with a risky asset strategy) can be viewed as increasing the shareholder option to default, and hence as a form of wealth transfer from policyholder to shareholder (as discussed above in Chapter 6.1). In this setting, the role of prudential solvency regulation can be viewed as limiting this transfer of wealth from policyholder to shareholder. This can be done by requiring the insurance firm to hold more capital as asset risk is increased. The increase in capital reduces the 'moneyness' of the option to default. This can (partially) offset the impact on the option value of the increase in asset volatility.

Such a model of shareholder incentives and regulatory control provides a potentially interesting perspective on the purpose and effect of prudential solvency regulation. More than this, it may even point to a way of *quantifying* prudential capital requirements. *As an alternative to a run-off or VaR definition of capital requirements, the capital requirement can be defined as the amount of capital required such that the default put option is limited to some specific level*. For example, we might define that level as 1% of the present value of the policyholder liability (valued when it is assumed to be default-free). This option valuation would refer, where available, to the market prices of traded options in the risks that drive the capital requirement.

Let us develop a very simple (and rather pathological) example to illustrate the dynamics of this approach at work. Suppose an insurer has a 10-year fixed liability cashflow of 100, and the basic reserve for this liability is assessed by discounting the cashflow at the 10-year risk-free rate. The 10-year risk-free rate is 2% (continuously compounded) and the basic reserve is therefore 81.9.

Instead of investing in the 10-year risk-free bond, however, the insurance firm decides to invest this reserve entirely in an equity index. To assess the capital requirement under the run-off approach, the probability distribution for the 10-year equity index value would be required. To assess this capital requirement under a 1-year VaR approach, the joint probability distribution for the 1-year equity index value and the 1-year change in the 9-year interest rate would be required (the 9-year interest rate is required in order to value the liability cashflow after 1 year, as required by the VaR calculation).

How would this capital be assessed using the default option approach? The shareholder option to default can be valued as a 10-year put option on the equity index with a strike of 100 and a spot value of 81.9. Further suppose the 10-year equity index put option at this strike has a Black-Scholes implied volatility of 16%. Then, in the absence of any additional capital, the shareholder option to default has a value of 16.4, which is 20.0% of the basic reserve. Now suppose additional capital will be held and will be invested in the matching risk-free bond. This capital requirement definition requires us to answer: how much additional capital is required to reduce the put option value to 1% of the basic reserve? The answer is 43.5. Investing this additional amount today in the risk-free bond reduces the strike of the 10-year option by 53.1 from 100 to 46.9. When the put option's strike is reduced to 46.9, its value is reduced from 16.4 to 0.82, i.e. 1% of the basic reserve. *So, this approach has assessed that a capital requirement of 46.9 results from investing the 81.9 of basic reserve in equities instead of the matching bond.*

This measure of solvency capital requirement has some distinctive features relative to the VaR approach used in Solvency II and the run-off approach that has been used widely in recent actuarial history. In particular, it offers two advantages over these other quantitative definitions for risk-sensitive capital requirements:

- *This definition of solvency capital makes no direct use of probabilities*. This is philosophically appealing to the social science anti-positivist who cannot subscribe to the idea that these probabilities can be estimated with reliable accuracy or, indeed, are particularly meaningful.

- The capital measure defines how much of the risky asset return should be passed over to the policyholder in the form of a (fixed) reduction in the premiums charged for the insurance policy (if the capital requirement is defined at the 1% level that would be commensurate with a 1% reduction in the premium charged). *This measure is therefore a more economically meaningful quantity than an arbitrary percentile* (or Conditional Tail Expectation) level.

However, such a definition is, of course, no panacea. Many of the financial market risks (and non-financial risks such as longevity risk) that are found on insurance balance sheets are not risks for which there are observable traded option prices. As a result, the option valuation exercise described above would need to use a valuation method that heavily relied on 'real-world' probability estimates to generate inferred values for these non-existent option prices. But actuaries already regularly make use of such assumptions in the valuation of contingent liabilities such as with-profit guarantees, so this does not present any significant new challenges beyond what is already required by the VaR approach.

Case Study (2) - The Advance Funding of Defined Benefit Pension Funds

The advance funding of private sector Defined Benefit (DB) pensions has been an established feature of employee remuneration in the UK since the latter part of the 19th century (though today it has reached the stage of managed terminal decline)[396]. Actuarial advice on the setting of contribution rates and the broader financial management of these pension funds has been one of the core areas of UK actuarial practice for over 100 years.

Much of this actuarial work is involved with the estimation of pension fund members' future mortality rates, salary increases, and so on as part of the projection of the long-term liability cashflows of the pension fund. As per Chapter 5.4, there is considerable scope for errors in the long-term prediction of these variables and the liability cashflows that depend on them. But the most historically contentious and consequential aspect of actuarial work in DB pensions lies not in the projection of liability cashflows *per se*, but in the way the projected liability cashflows are transformed into a liability 'valuation'. That is, in the language of Chapter 5.2, it is the choice of discount function that is used in the actuarial model to transform the projected cashflows into some form of present value metric that is the most interesting aspect of actuarial methodology in DB pensions.

This liability valuation can be used to perform at least a couple of tasks: it can determine what is commonly referred to as the current funding level of the pension fund - whether there is, in the actuary's professional opinion, sufficient assets in the pension fund today, given the pension fund's liabilities, asset strategy, strength of sponsor, and so on; and it can be used to determine the contribution rate that the actuary believes will be sufficient, when considered alongside the current assets of the fund, to meet future liability cashflows as they fall due, again given the fund's asset strategy.

How does the actuary determine whether the current funding level of the pension fund is adequate? A measure of the funding adequacy of a Defined Benefit pension fund should naturally follow from a basic starting rationale for the advance funding of the benefits: *in order to assess how much is enough, we must first be clear why there is a need for any at all*. Historically, a number of such rationales have been forwarded and discussed by the actuarial profession[397]. The debate about the primary purpose of advance funding of DB pensions has arguably never been adequately resolved within the UK actuarial profession. As early as 1911, three main candidate rationales had been documented and discussed in the actuarial literature: to provide security of benefits to the

[396] See Turnbull (2017), Chapter 6 for historical background.
[397] See Turnbull (2017), Chapter 6 and, in particular, the section 'Funding for What?' p. 254-265.

employee who had been promised long-term pension benefits by a private sector employer; to reduce the ultimate cost of providing the pension through successful investing of the advance contributions; to smooth the annual expense to the provider of the pension promise[398]. These are three very different reasons for advance funding, and they imply very different ways of determining the rate of advance funding and how (or, indeed, if) funding adequacy is measured.

The last of the three rationales discussed in the 1911 paper described above had already became the dominant perspective that drove actuarial methods in DB pensions at this time. This remained the case throughout most of the 20th century and arguably even to this day. That is, the fundamental actuarial assessment in DB pensions is the calculation of the contributions, expressed as a constant proportion of salary, that are required to provide a reasonable expectation that the pension promises can be paid in full as they fall due.

Pension actuaries have many subtle definitional variations on that theme, but the key point here is that this measure of the required contribution rate and related assessments of funding adequacy are assessed using long-term and subjective estimates of various socio-economic phenomena such as price inflation, salary inflation, risk-free interest rates, excess expected returns of risky assets, and longevity improvements. From the perspective of the sceptic of positivist methods in social science as articulated in this work, it is very hard to see how these assumptions can be estimated with any reliable accuracy. It may be countered that this does not really matter; that by continually updating our best estimates of these assumptions and revising our recommendations as a result, the finances of the DB pension fund can be steadily steered in the correct long-term direction. But once we recognise such non-stationarity and uncertainty in these phenomena, the entire perspective of estimating a long-term stable funding strategy surely starts to sound like an unsatisfactory answer to the wrong question.

This approach to assessing the financial health of DB pensions also ignores what is arguably the most important rationale for advance funding of DB pensions: to provide security today for the benefits that have so far been promised to the pension fund member. It is the risk that the sponsor proves unable or unwilling to meet their accrued pension promises that provides the rationale for putting pension fund assets in a legal trust structure distinct from the employer balance sheet. It is this risk that elevates actuarial advice on DB pension fund management beyond being just another part of the finance function of the sponsoring employer. If it could be known that a pension fund sponsor could never default on their pension promise, the need for advance funding of DB pensions would be vastly diminished. And, similarly, the particular level of funding of a pension fund at any given point in time would be a concern to no one but the sponsor.

Once, however, the risk that the sponsor may prove unable to meet the future cost of past accrued pension benefits, the security rationale for the advance funding of DB pensions is obvious. Moreover, a clear funding adequacy test logically follows from this advance funding rationale that is quite distinct from the long-term funding level and contribution rate estimate approaches described above: *from the member security perspective, the pension fund should have sufficient assets to fund the transfer of the accrued pension promises to a financially secure third-party at any time (including now, and, in particular, in the potential future circumstances that lead to the employer sponsor becoming financially distressed). Such an assessment of funding adequacy does not rely to the same degree as traditional actuarial pension funding assessments on long-term subjective estimates of socio-economic phenomena.*

[398] Manly (1911), p. 219; Turnbull (2017), p. 237 for further discussion.

Whilst it has never dominated pension actuarial thought, the argument for giving priority to the security rationale for advance funding is not a new idea. It certainly features in the pension actuarial literature of the 1970s and 1980s[399]. By the end of the 1980s, there was an increasing acceptance that the cost of securing the benefits with a third party should be considered in the actuarial analysis of DB pensions. Some leading pension actuaries proposed that this buyout cost ought to define a minimum funding level for the pension fund[400]. This type of thinking led to new UK government legislation in the mid-1990s that was intended to ensure that such a minimum funding level was achieved by private sector pension funds. But the 'Minimum Funding Requirement' was fudged and watered down to such a degree as to be largely ineffectual.

With the benefit of hindsight, this may be seen today as an opportunity missed. The unexpected falls in long-term interest rates and the lower future assets returns that are anticipated as a result means that the 'traditional' actuarial strategies for DB pension fund management have not fared well over the last 25 years. At the time of writing, the aggregate deficit of UK private sector pension funds on a 'buyout' basis is well over £500bn[401]. Of course, pension actuaries cannot be faulted for failing to predict the unprecedented falls in interest rates of the last 25 years. But this rather misses the point. If actuarial methods had placed greater focus on the security rationale, their methods would not have relied on these predictions to the same extent. The actuarial answers required by the security rationale do not require assumptions about the equity risk premium or the expected risk-free interest rate over the next 40 years. It is a simpler question that is more robust to deep uncertainty. Moreover, it is one that is better aligned to pension fund members' interests. If the financial management (contribution rate, funding adequacy, investment strategy) of DB pension funds had been primarily driven by the benefit security rationale over the last 25 years, the DB pension fund sector would be in materially better financial health than it is today.

What generalisations can we make from these examples?

Hopefully, the previous 100,000 or so words of epistemic scepticism have imbued the reader with a reticence to make sweeping universal generalisations on the basis of a couple observations. Nonetheless, the logic that has been applied in the development of these two case studies is fundamentally similar in some respects. The most obvious similarity between the two examples is that *they have attempted to resolve a 'traditional' actuarial question that is based on the subjective measurement of long-term expectations and / or probabilities into a question that is based on current economic values that make use of available market prices*. *It was argued in each case that this alternative calculation was more robust to deep uncertainty and more economically meaningful and informative as a risk management metric*.

It was noted above that the estimation of the current economic value of actuarial liability cashflows will rarely be completely straightforward. Liquid financial markets will usually only provide a sub-set of the information required to estimate these values. But some is better than none. For example, in the insurance capital case study, the observed equity index option price provided a crucial input into this capital assessment calculation. It removed the need for subjective estimates of the joint 1-year 99.5th percentile changes in the equity index and 9-year interest rate (VaR approach) or the 10-year equity index total return distribution (run-off approach). But if the investment strategy had been more exotic than an equity index – say, including a range of asset classes that are not publicly traded, such as real estate or private equity, it would be much harder to find relevant option prices to input into the capital assessment process. In this case, there will be little choice but to estimate

[399] Gilley (1972); McLeish (1983); McLeish and Stewart (1987).
[400] Thornton in McLeish and Stewart (1987), p. 155.
[401] Pension Protection Fund (2018)

these option prices using probabilistic assumptions together with other market prices of less direct relevance to the liability under consideration. Similarly, market prices will not be available for a range of typical insurance balance sheet risks such as longevity risk. But, in the absence of a liquid market for longevity risk, the modelling assumptions required for longevity risk in a market value-based analysis are not any more demanding than the assumptions required in a subjective probability-based modelling exercise.

We have forwarded two fundamental and distinct arguments in general favour of market value-based approaches to actuarial risk and capital assessment: these approaches will tend to be more *reliably accurate* and *more robust to deep uncertainty* than subjective probabilistic approaches; and they will provide information that is more *meaningful and useful in decision-making*. Let's briefly consider these two arguments further.

There is an epistemic argument for expecting a market value-based assessment to be more reliably accurate than a subjective probabilistic approach. Market prices in well-functioning markets are, arguably, valuable sources of information on consensus expectations. In the terminology of Chapter 1, market prices are set by the intersubjective probabilities of market participants. There is a strong epistemic argument for preferring intersubjective probabilities over an individual expert's subjective view. Indeed, approaches to 'structuring' expert opinion such as the Delphi method are concerned with the aggregation of expert opinions into consensus views. This is essentially an effort to move from an individual's subjective probabilities to consensus intersubjective probabilities. In well-functioning markets, market prices can do that job for us.

Moreover, this type of analysis is likely to provide more *meaningful information*, with more direct consequences and potential for use in decision-making. This is because the analysis based on market prices naturally provides more directly actionable insight. In the pension fund example, the metric identifies the means of managing the most tangible form of risk that the members of funded pension schemes face (that the pension fund has inadequate assets to meet the market cost of transferring the liabilities in the event of sponsor default). In the insurance capital case study, the economic value approach could be used to determine how to fairly price the insurance product. In the earlier discussion of Guaranteed Annuity Options, market value-based analysis provided key insights into forms of liability hedging strategy that could be implemented at that point in time.

The more traditional actuarial analysis provides a more 'passive' analysis. Their long-term nature provides less insight into immediate decision-making choices. The counterargument to this point is that a focus on market prices encourages too much focus on the short-term. There has been a very long historical tradition of actuaries being reluctant to make use of market prices because they are viewed as exhibiting excess short-term volatility that is felt to be irrational. And there is a significant stream of empirical economics research that can be used to support this perspective[402]. There seems to be little doubt that market prices are affected by a range of factors beyond an asset's expected cashflows and risk. And, in the language of probability, this significantly complicates the discovery of the intersubjective probabilities that are implied by market prices. But not all actuarial clients have the luxury of dismissing irrational market prices. Returning to the pension fund example, when a sponsor defaults, the value of the pension fund asset portfolio and cost of liability buyout are what the market says they are, not what the pension actuary believes they ought to be.

[402] Some of the classic papers in this field include Shiller (1981), Fama and French (1988) and Roll (1988). See also Cochrane (2005), Part IV.

Finally, it should be emphasised that the above perspective is not an argument for exclusivity: this discussion assumes a context where the actuary can use models and their metrics in an interpretivist rather than positivist way. In that setting, a range of diverse analyses (market value-based and subjective probabilistic) can provide an array of useful insights that the actuary can make use of in their professional work. The thrust of the above discussion is that, despite some movement in this direction over the last twenty years or so, market value-based analyses remain under-utilised in actuarial science, and they remain an untapped source of profound insight for actuarial science when used well.

You may or not agree with the specifics of these illustrative examples. They are complex topics that have been treated quite superficially above. But hopefully the consistent way of thinking that runs through them is illuminating, or at least thought-provoking. Chapters 6.1 and 6.2 can be summarised as follows:

- *Actuarial financial institutions should eschew the accumulation of forms of risk exposure that are subject to deep uncertainty.*
- *In the absence of such deep uncertainty, risk measurement is more straightforward. In that context, the actuarial role will be concerned with the efficient and equitable pooling of diversifiable risks.*
- *Where deep uncertainties are accumulated, the use of available market prices will be useful both as a means of inferring intersubjective probabilities for the future behaviour of these phenomena; and as the basis for actionable insight on relevant risk management choices.*

6.3 Data science and actuarial science

We conclude our review of the methodology of actuarial science with a discussion of the emerging methods of data science and their potential implications for the future of actuarial science. The discussion begins with some basic background on the historical development of data science and its notable features; we then provide a high-level overview of some of the most popular methods of data science, and discuss some of their methodological features; finally, we consider how this description of data science fits with the characterisation of actuarial science developed above, and what this suggests for the future methods of actuarial science and how they may make use of the techniques of data science.

Data science: some historical background and the basic ideas

The world has been in the throes of a ‘digital revolution’ since the electronic integrated circuit was first developed in the 1950s. The first and most direct consequence of this revolution has been a *huge increase in computer processing power* and speed: computation speed increased by a factor of a million between the 1950s and 2015[403]. There is a second key dimension to the digital revolution: the increase in the *quantity of data* that is available about the world around us. Today we live in a world that is continuously generating and storing quantifiable data about the physical world, and, increasingly, about human behaviour: countless networked wireless sensors continuously make quantitative recordings of all sorts of natural and human phenomena; website browsing histories grow continuously; and so on. And the cost of storing this data has shrunk enormously in recent decades – the cost of storing a gigabyte of data on a hard drive has fallen from close to $1m in the 1980s to 3 cents by 2014[404].

---

[403] Efron and Hastie (2016), p. xv.

[404] Siegel (2016), p. 112.

Over the last twenty years or so, a field of study has emerged that is interested in how this data, which may be generated incidentally as the by-product of some other intended purpose (sometimes referred to as 'found data'), can be put to use in the identification of patterns and relationships that can facilitate useful and reliable predictions about the future. This specific field of data science is sometimes referred to as predictive analytics. It has been applied to make predictions for just about anything on which relevant data can be found (future crime rates; the ranking of suspected fraudulent financial transactions; consumer spending patterns; infectious disease outbreaks; etc.).

The term 'Big Data' was first used in the early 2000s in the field of astronomy when new digital instrumentation generated a previously unprecedented amount of quantitative observational data. But in more recent years much of the interest in big data has migrated to the study of human behaviour and social phenomena. To date, however, the existence of this data has not revolutionised the empirical methods of academic research in the social sciences (which are still dominated by empirical techniques such as the Randomised Controlled Trials method discussed in Chapter 3). The main applications of big data in the analysis of human behaviour in recent years has not been for the purposes of the advancement of scientific knowledge in the social sciences, but to identify commercially valuable predictions about the tendencies of individual people. This distinct interest in *prediction*, as opposed to the more traditional objectives of statistical analysis of providing explanation or description, will be discussed further below. But put simply, the analytical focus of big data is on predicting what happens next, rather than delivering insight or explanation into why or how.

The world of data science, despite being described above as the product of a form of technological revolution, can no doubt be placed within the long-existing domain of statistical inference: after all, it unambiguously involves using historical data to make inferences about the likelihood of particular things happening the future. But the above brief outline already provides some indication that the sort of modelling and analytical processes that are applied in data science may be quite distinct to those used in classical statistics. In particular, in big data applications:

- The *sheer size of the data samples* that arise may be of a different order to classical statistics – sample sizes may be measured in the hundreds of millions or more (whereas it is not uncommon for classical statistics to work with data samples measured in the hundreds or even less);
- The *number of dependent variables* that are potentially relevant may be measured in the hundreds or thousands (whereas the methods of classical statistics usually work well up to around 20);
- And the *emphasis on predictive performance* rather than explanation or description is different from the usual emphasis of classical statistical inference (for example, predictive modelling may not be very interested in ideas that are central to conventional statistical modelling such as the statistical significance of a given variable or hypothesis testing).

This distinct flavour of statistical inferential method has its own name. The analysis of big data using computerised prediction methods falls under the general term of *machine learning.* It is a relatively young and fast-evolving field. Its essential structure, however, is straightforward:

- A set of data exists (this data is sometimes called *learning data*, *training data* or *training examples*) that consists of *n* observations;
- Each observation can be considered as a pair of data elements. One element is the *response variable* (the property we are interested in predicting in the future) and the other element

consists of a *vector of predictors or covariates* (the properties that may be useful in predicting the behaviour of the response variable)[405];

- A machine learning algorithm is applied to the data to generate a model or rule that can make predictions for the value of the response variable when presented with any vector of predictors (there are many different types of algorithm of varying degrees of complexity that may be used here);
- Predictive performance is assessed using out-of-sample *validation* testing, i.e. by comparing the response variable results produced by the model with the results from a data set that was not used in the model fitting.

n = all?

Some big data 'evangelists' have argued that the pervasiveness of big data turns the basic statistical concept of a sample into an anachronism: 'the need for sampling is an artefact of a period of information scarcity, a product of the natural constraints on interacting with information in an analogue era'[406].

We have, it is argued by these writers, reached the age of *n = all*. This is a highly aspirational perspective that is in danger of fundamentally missing the point. On the face of it, an *n = all* world has no need for any form of statistical inference, or indeed any form of inductive inference. After all, if we have already observed the entire population and all its characteristics, there is nothing left to infer about it. Computing power has indeed led to some such situations. Chess end-games provide a good example[407]. With the aid of the vast computation that is possible with today's processing power, it has been possible to tabulate all possible moves when there are 6 or fewer pieces left on the board (even with today's vast computational processing capabilities, the scale of complexity of chess is currently too great to do this when more than 6 pieces are left on the board). The analysis of all possible moves has permitted the perfect strategy to be exactly identified in any such scenario. No inference or estimation is necessary as the answers have been deduced and exhaustively tabulated.

It would, however, be a highly mis-leading over-simplification to conclude that the chess end-game analogy is relevant to the prediction of future human social behaviour. Irrespective of the size of our data survey, the basic purpose of this data analysis remains fundamentally inferential: it is *to use modelling techniques to make inductive inferences from the observed to the unobserved*. In the context of predictive analytics, the observed lies in the past and the unobserved lies in the future. Predictive modelling is an exercise in induction, and it suffers from all the usual challenges that make inductive inference so fundamentally problematic, especially for social phenomena.

There can be no doubt that the vast volumes of previously untapped data that are offered by the digital revolution can be capable of providing many new and powerful insights. But inductive inference is intrinsically fallible, irrespective of the volume of our observational data. Its fallibility lies in the necessity for an assumption of a form of uniformity in the behaviour of the data over time (past, present and future). The data is always historical. For history to be a guide to the future, the future must be like the past. This may be a reasonable assumption for the behaviour of sub-atomic particles, but is less obviously robust for, say, the internet browsing tendencies of teenagers and

---

[405] Technically, the pre-identification in the data of predictor variables and response variables is called supervised learning. The antonym, unsupervised learning, is where the data set merely consists of unstructured, unspecified data. Our discussion will focus on the supervised learning setting.

[406] Mayer-Schonberger and Cukier (2017), p. 13.

[407] This example is given in Mayer-Schonberger and Cukier (2017), p. 36.

how they correlate with their spending habits. The data in big data studies of human behaviour may become obsolete quite quickly as behaviours and technologies change in inherently *unpredictable* ways. The shelf-life of a predictive model may therefore be very short, and its results always fallible.

The Google Flu Trends story provides the most famous, or notorious, example of the fallibility of the analysis of big data. In the mid-2000s, a machine-learning algorithm was applied to data on vast numbers of individuals' web browsing histories in order to generate predictions of flu infections rates in different geographical areas. This algorithm generated some impressive results in 2008 when it accurately predicted where flu outbreaks would occur. And it did so around one week faster than the US government's Centers for Diseases Control and Prevention were able to with their more conventional (and expensive) methods based on the collation of medical reports. But a few years later Google's algorithm generated a substantial false positive: it predicted a major flu epidemic that never materialised. And because the algorithm has no causal structure (it is focused on predictive performance only), no one can be quite sure exactly why it failed. But the fundamental cause of the failure is inescapable: people's internet browsing habits had, for some reason or other, changed in some important way that the algorithm had not predicted. It is hard to think of a more non-stationary environment than the way people use the internet. This example highlights a fragility that is common to all inferential techniques: the assumption of some form of uniformity in human behaviour over time. The essential importance of this assumption does not diminish as n increases, even when it reaches the heady heights of 'all'.

Statistics and data science

It was noted above that big data analytics and the machine learning algorithms that it employs are usually focused on *predictive modelling*. This might sound like an overly obvious statement – isn't all statistical modelling in some sense used for prediction? The answer isn't as straightforward as you may think. 'Traditional' statistical modelling has arguably been more concerned with providing descriptions of or a form of explanation for the joint behaviour of some phenomena, rather than prediction *per se*. When a medical researcher uses a statistical model to interpret the results of a Randomized Controlled Trial, they are testing some form of pre-specified hypothesis that has postulated a form of relationship between some phenomena of interest. Predictive modelling usually doesn't start with any pre-specified hypothesis and doesn't concern itself directly with whether a particular variable is significant in influencing the behaviour of the variable of interest. The process is simply focused on finding the algorithm that produces the best predictive performance (as measured by out-of-sample validation testing). What happens inside that process is treated largely as an incidental and noninterpretable black box, rather than as a hypothetical specification of causal probabilistic relationships[408].

The classical statistical methods established in the 1920s and '30s by Fisher, Neyman and Pearson used logic and mathematical theory to establish estimation or inferential methods that delivered, in some well-defined sense, optimal inferential solutions to well-specified problems under well-specified conditions (a sample's maximum likelihood estimate of a parameter value of a normal distribution, for example). This classical statistical theory was built using mathematically tractable parametric probability models where the dimension of the parameter vector was usually less than 20. Computational power was limited, and the methods' implementation relied on their mathematical tractability.

[408] See Shmueli (2010) for a more extensive discussion of the distinction between predictive and explanatory statistical modelling.

The increasing availability of computational power that began with electronic computation in the 1950s led to many new developments in statistical ideas, methods and techniques over the second half of the 20th century which were increasingly computationally intensive (examples include empirical Bayesian methods and the bootstrap). Such approaches can demand a thousand times more computation than the pre-war classical methods, but all or most of these 20th century innovations can still be quite naturally located within the framework of classical statistical theory. That is, they are still attempts to answer the questions defined by the classical theory.

The 21st century world of machine learning contrasts more sharply with classical statistical theory. Little or no inferential parameter optimality theory exists today for machine learning algorithms. There are many different machine learning algorithms, but they share an essentially similar form: an optimisation process that works by computational trial-and-error on a vast computational scale. The data scientist is focused on what works well for a specific and often messy and very high-dimensional problem. They are less interested in a theoretical justification for the method or its performance. In place of statistical theory, data scientists will commonly evaluate the performance of a new machine learning algorithm by benchmarking its predictive performance against well-known publicly available large data sets. This can make statisticians uneasy. To quote from the influential Efron and Hastie (2016): "the very large data sets available from the internet have encouraged a disregard for inferential justification of any type. This can be dangerous. The heterogeneous nature of "found" data makes statistical principles of analysis more, not less, relevant."[409]

Classical statistical methods were, in the main, first developed by statistical theorists, and then put into action by data practitioners in a very wide range of fields (including the actuarial one, of course). The chronology of 21st century machine learning development has been reversed: new predictive algorithms are more likely to make their debut appearance in the computer science literature than the statistics literature. Innovative predictive algorithms are created in an experimental and *ad hoc* way at the coalface of high-dimensional big data applications with vast computational power on tap. Those algorithms that are identified as apparently working well may be heralded with much fanfare and hope. The statisticians currently tend to follow in their wake, working out how well they really work and why; and when and how they fit into their existing theoretical frameworks.

Machine learning methods

Machine learning algorithms can generally be viewed as a form of regression method. These regression methods, however, may look like only distant cousins of standard parametric linear regression or logistic regression models. The machine learning methods have been evolved by experience to perform well with very large data sets where the potential number of regression variables might number in the hundreds or even thousands. They do this by taking advantage of cheap computation and paying less attention to mathematical tractability or inferential logic.

Let's briefly consider at a high level some of the basic forms of machine learning algorithm and how they are fitted to data. The regression tree (sometimes called decision tree) method is perhaps the most standard, basic machine learning method. Its most common implementation method, Classification and Regression Trees (CART), is based on a method published in 1984[410]. Regression trees deliver a non-parametric, piecewise recursive partitioning of the data into homogenous groups. It is a classic example of computation replacing mathematical formulae. A number of

[409] Efron and Hastie (2016), p. 230.
[410] Breiman, Friedman, Stone and Olshen (1984). Siegel (2016), p. 178.

algorithmic rules exist that perform this data partitioning, with each recursive step splitting the data into two categories that are as distinct from each other as possible (or, equivalently, as homogenous as possible within each group)[411]. The regression tree method has been shown to be less efficient and accurate than some other more sophisticated machine learning algorithms, but it is very easy to computationally implement and requires virtually no assumptions about the particular probability model that is driving the phenomena, so it can quickly produce reasonable results for a very wide range of data and predictive problems.

One of the fundamental technical challenges in the implementation of predictive models such as regression trees is avoiding 'over-fitting'. Ultimately, any model can perfectly fit a set of data given enough parameters or algorithmic steps. But data contains both noise and signal, and the predictive modelling process aims to fit only to the latter. Statisticians conventionally manage the over-fitting problem by adjusting their fitting processes to penalise the number and / or magnitude of fitted parameters (the Akaike Information Criterion (AIC), which itself is a relatively modern development, first published in 1973, is one example of many such approaches that have been developed). The statistician would then use out-of-sample- validation as a way of independently checking the quality of the fitted model.

Data scientists tend to use validation techniques more directly in the model selection / fitting process, and as their primary means of avoiding over-fitting. Validation means using another data set (distinct from the training set) to assess the discrepancy between the model's predictions and the responses in the validation data. Validations can be performed on any number of candidate models, and the model that generates the lowest average discrepancy may then be selected as the preferred model. Some statisticians may argue that the point of validation testing is not to select a model, but to validate that the selected model works. This is perhaps another example of data science foregoing statistical formality for a brute-force assumption-free general method that usually works. More traditional statistical model selection methods such as the AIC can provide more accurate assessment of models, but such an approach rely on probability modelling assumptions that may or may not pertain to a given problem[412].

Out-of-sample validation means sacrificing some sample data by removing it from the training process and dedicating it to the validation exercise. In the context of the huge data samples available in Big Data problems, this sacrifice may not appear especially costly. However, the vast number of predictor variates that may be considered in these exercises demand a lot of training data in order to have a chance of predictive accuracy. So, methods that can limit this sacrifice of fitting data are valued. *Cross-validation* is one such approach. Suppose we have a number n of training data pairs. The usual validation approach would require removing 20% or 30% of the training data and quarantining it for validation testing. The cross-validation approach avoids this sacrifice. With the cross-validation method, n validation results are produced by re-training the model n times, each time with a single data point removed (variations exist where more than one point is removed in each validation run). The single removed point is independent of the training set and so can be used as a validation case. This produces n validation results for the model, and this process can be run for each candidate model. This clearly involves enormous computation. But, once again, that is no major problem for modern computing technology. In regression trees methods, cross-validation approaches are usually used to determine at what step to stop branching out further[413].

---

[411] See Efron and Hastie (2016), p. 126 and p. 130.
[412] Efron and Hastie (2016), p. 227.
[413] Efron and Hastie (2016), p. 126.

In the 21st century, the increased understanding of the predictive power of *ensemble modelling* has been a major theme in the development of machine learning algorithms. The basic idea of ensemble modelling is not especially mathematically sophisticated: ensemble modelling involves constructing a predictive model as an aggregate of a number of independently fitted predictive models. Ensemble models have been found to perform consistently better than a single predictive model that is based on the same overall data set. Once again, the computational requirement is much greater, but today computation is cheap.

The most prominent ensemble modelling example is the method known as *random forests*, which was introduced in 2001[414]*.* The random forests method is an ensemble method for regression trees. Random forests have produced some of the best-performing predictive machine learning algorithms. They do so by using a method that is again very general and that can be implemented virtually automatically without specific probability modelling assumptions. The basic idea of the random forest is to re-create the underlying data set many (i.e. thousands of) times using the bootstrapping re-sampling method. A regression tree is then fitted to each of these data sets. And the random forest is then constructed as the average of the fits of the regression trees. Cross-validation results are produced almost automatically as a result of the random forest algorithm. In each bootstrapping data re-sampling process, some of the data points will be left out of the re-constructed data set. These data points are therefore independent of the fit that has been produced for that data set and can be used as validation data.

The random forest algorithm is easy to implement, in the sense that it does not require many probability distribution assumptions or fine tuning. But its non-parametric ensemble output structure can make it difficult to interpret, especially when the dimension of the predictor variables is high. Many other machine learning algorithms exist that can offer a different trade-off in terms of predictive performance, ease of implementation and transparency of output. These include boosting, neural networks (sometimes referred to as deep learning), support-vector machines and kernel methods. All of these can be viewed as some form or another of regression model, with an implementation emphasis on its use in prediction.

Data science and hypothesis testing: the 'multiple comparisons trap'

It was noted above that machine learning and big data analytics tended to place less emphasis on formal hypothesis and significance testing than is the case in conventional statistics. And you might expect that the vast scale of data used in big data exercises would make formal statistical significance testing nearly redundant. After all, the standard errors in the estimates from huge samples should be extremely small: if a relationship between two variables is implied by so much data, it surely can't have arisen from random sampling error.

But there is an important complication that commonly arises when considering this logic in the context of big data analysis: the process does not test one specific relationship of interest, but algorithmically searches out any relationship that can be found in the data. This is a quite different context to the usual hypothesis testing and significance testing situation. There, we may start with the null hypothesis that one specific relationship does or does not exist, and then test to see if the evidence in the data is strong enough to reject that hypothesis. In the big data context, we may similarly start with the null hypothesis that there are no relationships in the data, but then test many hundreds or thousands or even hundreds of thousands of different possible relationships. Suppose we use a 1% significance level. It is virtually inevitable, by the basic logic of probability, that some relationships will be identified as statistically significant in a set of completely random data – if we

[414] Breiman (2001)

test 5000 relationships at the 1% level, on average we will find 50 false positives. In machine learning, data scientists sometimes refer to this as the *multiple comparisons trap*[415]. Statisticians recognise this type of problem more generally in the form of *large-scale hypothesis testing*.

How can the multiple comparisons trap be avoided? The simplest solution is to make the probability level for hypothesis or significance testing more severe. For example, it could be shifted to 0.1% or even 0.05% instead of the usual 1% or 5% used in classical statistical significance testing. The *Bonferroni bound* was one of the first approaches of this kind proposed in the statistics literature. It was developed in the context of large-scale hypothesis testing in the late 1950s and early 1960s[416]. It strengthens the significance level by a factor of 1/N, where N is the number of hypotheses being simultaneously tested. Today this is regarded as a highly conservative approach. Statisticians have developed a field called False Discovery Rate Control that provides a number of more sophisticated and complex rules for setting significance levels for large-scale hypothesis tests. These rules tend to produce less conservative significance level adjustments than the Bonferroni bound.[417]

The drawback with the solution of making the significance level more severe is, of course, that by reducing the probability of false positives the probability of generating false negatives is inevitably increased. But if sample sizes of vast sizes are available (that meet the usual statistical testing criteria of being independent and identically distributed), then statistical tests of a high power may still be obtained, even with these unusually severe significance levels.

There is a fundamental point lurking in this discussion – when searching for relationships in an entirely inductive and unstructured way, without any underlying theory about the relationships that hypothetically may exist within the data, then the quantity of well-behaved data required to reliably say anything is much larger than when we have established a prior reason to search for and test a specific relationship. This brings us to the next topic in this brief discussion of the methodology of machine learning and data science.

Data science and the scientific method: correlation, not causation

We noted above that big data analysis and its machine learning algorithms typically do not seek to provide causal explanations for why a given identified relationship holds (or indeed even *if* it holds with any particular level of statistical confidence). There is nothing to theoretically prohibit the researcher from specifying the predictor variables that will be used in the machine learning algorithm, perhaps in the hope of establishing explanatory relationships. But, more typically, the model simply goes with whatever the algorithm identifies from the data as working as a predictor. The dependency relationships that are identified as important (and, conversely, the relationships that can safely be assumed to be irrelevant) are of little direct interest. Instead, the focus is on the predictors' collective role as a means to the ends of prediction.

It is worth reflecting further on how this use of statistical modelling contrasts with the conventional application of statistical modelling in the scientific method as discussed in Chapters 2 and 3. We saw in Chapter 2.1 how Popper and the logical empiricists were able to fit probabilistic hypotheses into the hypothetico-deductive scientific method and its essential logic of falsification. To recap, when probabilistic statements are included within the hypotheses, these can be subjected to hypothesis testing, where the statistical significance of the hypothesised relationships is assessed. Depending on your philosophical persuasion (realist or instrumentalist), you may view these relationships as

---

[415] See Siegel (2016), p. 137-42.
[416] Dunn (1961)
[417] See Efron and Hastie (2016), Chapter 15, for a technical overview of the large-scale hypothesis testing and False Discovery Rate control topics.

explanatory or descriptive, but this difference is 'merely' philosophical, in the sense that it has little direct implication for the execution of the scientific method.

As noted above, predictive modelling does not do statistics in this way. It does not specify structural or causal relationships for statistical evaluation. It simply searches everywhere (perhaps amongst thousands of variables) without prejudice or prior assumptions for whatever relationships it can find that are useful in out-of-sample prediction. In some cases, it may not even be possible to access what relationships have been identified by the modelling algorithm. The potential complexity of machine learning algorithms can create a 'black box' characteristic where it is not possible to trace the logical steps that have been used in the derivation of the model. This 'non-interpretability' or lack of 'explainability', as it is called in the field of artificial intelligence, may create challenges for the broader validation of the model and the decisions based upon its output.

Some big data evangelists have rejoiced in this escape from the delusion that knowledge of causation is attainable: 'The ideal of identifying causal mechanisms is a self-congratulatory illusion; big data overturns this.'[418] Some have even argued that predictive analytics will herald the end of the hypothetico-deductive scientific method altogether[419]. Others have argued that predictive modelling can be a useful complement to the explanatory statistical modelling that has been used in the modern scientific method for the last 50 or 70 years or so. In particular, it has been argued[420] that predictive modelling may be useful in finding new empirical relationships that can then provide the impetus for the development of theories according to the hypothetico-deductive method. From here we might recall Chapter 2.3 and its discussion of empirical generalisations such as Boyle's Law. Such laws were merely noted recurring empirical relationships, stated without any argument for *why* such a relationship exists. The *why*, such as the Ideal gas law, that deduces the empirically-observed relationship (and other testable relationships so as not to be merely *ad hoc*) from some prior basic premises, may then follow.

It seems, in principle, quite feasible that predictive modelling can find new empirical generalisations amongst observable phenomena. This, obviously, requires the predictive modelling to deliver its output in terms of transparent and interpretable relationships between the variables in the data, and not just in its usually currency of out-of-sample predictive power. And it requires critical thought to manage the potential problem of false positive discovery rates that will likely arise from such processes (Chapter 2.1's discussion of the 'replication crisis' will be particularly germane in the context of vast indiscriminate searches of unstructured data).

Data science and actuarial science

Actuarial science, of course, has a long history of working with large data sets to generate probabilistic forecasts of the future. An excellent early historical example of actuarial work with very large data sets can be found in the pooling of life office mortality experience. This, in data science parlance, might be thought of as a 19th century attempt at *n = all* data analysis: the ambition was to combine life assurers' mortality experience data to provide an analysis of all life assurance experience. In 1881, American actuaries published the Thirty American Offices' mortality tables, which used the mortality experience data of over 1 million life assurance policies over the 30-year period between 1844 and 1874[421].

---

418 Mayer-Schonberger and Cukier (2017), p. 18.
419 Mayer-Schonberger and Cukier (2017), p. 70.
420 For example, see Shmueli (2010), Section 4.1.
421 Turnbull (2017), p. 86.

The calculations employed in the data analysis of the 1870s American actuary, however, profoundly differ from those used by the machine learning algorithms above. And the mortality modelling methods of the 2020 American (or indeed any other) mortality actuary will probably share more in common with the 1870s actuary than with the machine learning predictive analytics described above. Nonetheless, data science does share some significant methodological similarities with actuarial science. Chapter 5 noted how philosophers of science characterised actuarial science using terms such as empirical, applied, local, approximate and bottom-up. These terms are also quite applicable to data science. Indeed, the above discussion of machine learning methods might suggest that terms such as these apply to data science even more emphatically than they do to actuarial science. This suggests the techniques of data science might fit quite naturally into the statistical / predictive toolkit of the actuary.

As noted above, machine learning algorithms can be cast as some form of regression method. As has been pointed out in recent actuarial literature[422], many actuarial models can also be cast as a form of regression model. Non-life insurance pricing formulae, mortality forecasting models, and curve fitting of asset and liability valuation projections are all clear examples. This makes the scope of application for machine learning within actuarial science potentially very wide. Recent studies have begun the process of evaluating the impact that data science methods can have on actuarial problems. These impacts can be considered in the two dimensions of the digital revolution: the impact that machine learning algorithms can have when these computationally intensive techniques replace the traditional statistical techniques conventionally applied in actuarial science; and the impact of the increase in the type and volume of potentially relevant data that is available for use in actuarial problems.

There are a number of studies in the contemporary actuarial literature[423] that consider only the first of these dimensions – that is, where machine learning algorithms replace conventional actuarial probability models and are applied to the usual actuarial data. For example, a recent study compared the performance of neural networks with that of the Lee-Carter mortality model in long-term mortality forecasting[424]. Another recent study showed how various machine learning algorithms perform as an alternative to generalized linear models in motor insurance claims frequency modelling[425]. However, in both of these examples, machine learning algorithms did not provide a meaningful improvement in out-of-sample predictive errors relative to the contemporary versions of the actuarial probability models. This seems fairly intuitive: these problems are not big data problems (at least not in the context of these studies). The mortality study used 59 annual data points for 80 age groups. The predictor variate is of low dimension in both cases – there are 9 predictor variables in the motor insurance example, and 2 in the mortality example. It is not obvious what machine learning algorithms can do with this data that conventional probability modelling methods cannot.

The situation is much more interesting when the nature of the data that is available for actuarial use is fundamentally different to historical actuarial norms. New forms of data with potentially significant value to insurance underwriting and product design has undoubtedly started to emerge in major scale in recent years. For example, the idea of the 'Quantified Self', which refers to the cultural phenomenon of individuals self-tracking various aspects of their physical and medical

---

422 Richman (2018), p. 5.

423 See Richman (2018) for a comprehensive survey of the contemporary actuarial literature on the application of machine learning techniques.

424 Hainaut (2017)

425 Noll, Salzmann and Wuthrich (2018)

experience, often using wearable technology, has the potential to revolutionise life and health underwriting. But perhaps the most-developed example of this data revolution for insurers can today be found in motor insurance in the form of the telematic data that can be provided by sensors in motor vehicles (and parallels are likely with other forms of non-life insurance that are touched by the 'Internet of Things'). A range of vehicle sensors may be installed in cars. But just one single GPS sensor delivering second-by-second location data may provide information of transformational value for underwriting purposes. Why? Because such high frequency location data can allow the driver's speed and acceleration patterns to be observed.

Traditional actuarial techniques arguably do not provide an obvious answer for the most effective handling of second-by-second data feeds. Recent actuarial literature has investigated how machine learning algorithms can use this speed / accuracy data to classify driving styles[426]. This classification can, for example, allow a 'safe driving index' to be measured for individual drivers. Research on the implications of this telematics data for motor insurance pricing is still in its early stages. But this study has suggested that the inclusion of these driving style classifications, derived solely from telematics data, significantly improves out-of-sample prediction for motor insurance claims frequency, and may even be more important than 'classical' motor insurance rating factors such as driver's age.

Again, the methodological conclusion here is intuitive. When the data available to the actuary becomes much bigger (both in number of observations and number of dependent variables), it is likely we will be entering a domain that is ill-suited to traditional statistical techniques and that is instead the natural hunting ground of machine learning algorithms. In such circumstances, these techniques may be indispensable tools in the development, pricing and ongoing management of insurance products. In short, the availability of this type of data is likely to revolutionise insurance pricing, and it is likely that machine learning techniques will be necessary to make the most use of it. As always, the actuary's role will require judgment to be exercised in interpreting what these models can be used to reliably infer, and how this can be most effectively put to commercial use (and within ethical boundaries).

The Actuarial Profession and its Science: A Sociological Perspective

Most professions, most of the time, have professional jurisdictions whose boundaries are in a constant state of flux. Professions are generally about the trusted delivery of expert services to human problems. New theories and technologies, exogenous factors such as trends in societal demands and so on change the nature of these tasks and the level of demand for them, thereby creating opportunities and threats in which professions compete with each other to adapt, survive and thrive. This interprofessional competition is a fundamental driver of how professions evolve through time. For many years, the actuarial profession had an unusually stable professional jurisdiction. In the late 1980s, one of the world's leading sociologists of the professions noted that the actuarial profession was 'atypical' in its lack of interest in expanding its professional jurisdiction[427].

Over the last twenty years, however, this picture has altered somewhat radically for the actuarial profession. Since then, much of the profession's 'heartland' legally-protected jurisdictions have been in a form of existential crisis: Defined Benefit pension funds are now in a state of managed terminal decline; with-profit funds even more so. At the same time, some of the financial theory and knowledge that has been developed by economists over the second half of the twentieth century

---

426 Wuthrich (2017)

427 Abbott (1988), p. 71.

has increasingly led to the economics profession challenging some of the traditional methods of actuarial practice by reducing apparently opaque actuarial problems into well-understood tasks amenable to economic theory. The actuarial profession's reaction has sometimes been caught between indecision on whether to absorb this knowledge into its own methods and systems or to reject the knowledge as irrelevant to the profession's tasks. Unsurprisingly, these circumstances have also led the profession to place new energy on the identification of new tasks for its skills. 'Wider fields' has been a continuing preoccupation of the profession's leadership since the start of the twenty-first century, if not before.

Abbott argued that it was what he called a profession's 'abstract knowledge systems' that allowed professions to adapt and claim new professional jurisdictions. Indeed, he argued, the presence of such abstract knowledge systems are an important part of what differentiates a profession from other occupational groups: 'Many occupations fight for turf, but only professions expand their cognitive dominion by using abstract knowledge to annex new areas, to define them as their own proper work'.[428]

'Abstract' here means knowledge that is transferable, that it can be applied to a range of different specific problems. Such knowledge provides professions with the power to identify new domains and applications for their skills, and to compete to perform tasks that are currently performed by other professions, and to help explain why they are the right experts for their existing jurisdictions.

Data science and the actuarial profession

Does this perspective on the evolution of professions help us understand how the increasing relevance of data science might impact on the future professional activities of actuaries? It is natural that some actuaries will specialise in research on data science techniques and their application to questions within the professional actuarial domain. Some actuaries who specialise in data science may also venture outside of the actuarial domain and find other applications for their data science expertise.

It is likely to be equally important for the profession to work collaboratively with non-actuarial data scientists in the application of data science techniques to actuarial problems. Many professions can seem insular and impenetrable to outsiders. The actuarial profession has doubtless been historically subject to such claims. It is easy for collaboration between any two professions or disciplines to descend into little more than jurisdictional competition and rivalry.

Actuarial science is arguably unusual in the extent to which it makes use of technical knowledge that is independently developed by other professions and disciplines (such as statistics and economics). It is therefore likely to be crucial to the long-term quality of the technical content of professional actuarial work that the profession can find positive ways of collaborating with other technical disciplines in order to successfully incorporate the new technical knowledge developed outside the profession into actuarial science. The profession's long-term historical record in this respect, in areas such as economics, has arguably been somewhat mixed. But the above discussion has highlighted how the technical content of data science fits into actuarial science quite naturally. From this perspective, the interesting historical case of architects and structural engineers can perhaps provide an inspiring model of how technological transformation can give rise to new technical professions which successfully work alongside long-established learned professions in mutual self-interest and in the interests of their clients.

[428] Abbott (1988), p. 102.

Appendix – The historical development of economics

Economic thought can be found, to a greater or lesser degree, throughout all of the major philosophical epochs since the Ancient Greeks. There is some treatment of topics of economics in the writings of Plato, but it is the work of Aristotle that earns him the accolade of being 'the first analytical economist'[429]. He wrote most notably on the analysis of exchange and the role for money that it created.

The Romans left behind very little by way of philosophical writings about economic ideas. The monastic philosophy of the medieval times also did not particularly focus on economic thought. The ideas of this era tended to have a strong moral and religious dimension (such as the prohibition of taking interest from lending) rather than a positive focus on developing economic organisation for the purposes of improved production or distribution of wealth. These meagre developments in economic thought reflected the prevailing economic reality of society. The very limited opportunities for profitable investment in the medieval economy meant that money lending tended to be an activity that focused on the exploitation of the needy. The relatively simple forms of economic structures and relationships that prevailed provided little inspiration for developments in economic theory.

Petty, Hume and Smith

With the middle ages coming to an end, the fifteenth, sixteenth and seventeenth centuries saw economic society develop rapidly. Revolutions in farming methods increased agricultural productivity and rendered long-standing feudal relationships unsustainable. Maritime adventure brought the beginnings of international trade. Economic thought started to reflect these new circumstances. Most notably, Sir William Petty's economic writings of the 1680s discussed the productivity benefits that could accrue from the division of labour and the role this played in the development of large towns[430].

It was the period of the Enlightenment that saw economics emerge as an established field of study. David Hume offered some economic thinking within his great philosophical works, such as his chapter *On Property and Prices* in Book II of *A Treatise of Human Nature*, first published in 1739. His essay *On Money*[431] is now regarded as the definitive 18th century version of the quantity theory of money (though not as the *first* quantity theory of money, which is usually credited to the Frenchman Jean Bodin in his book of 1569[432]).

But it was Adam Smith's *An Inquiry into the Nature and the Causes of the Wealth of Nations*, published in 1776 in the midst of the beginnings of the Industrial Revolution, that proved to be the most influential and durable contribution to economic thought of the eighteenth century. It is often taken as the starting point for modern economics and Smith has been said to be 'universally acknowledged as the founder of classical political economy'[433]. Smith, like Hume, was a Scots philosopher, but Smith fully 'converted' to the study of economics, and in so doing became the first academic economist. Few of the ideas in *Wealth of Nations* were truly original. But his compelling articulation of these ideas as part of a coherent structure of economic thought had an immediate

---

[429] Roll (1992), p. 20.
[430] Petty (1899). This work is also notable in actuarial history for its inclusion of John Graunt's mortality table, first published in 1662.
[431] Hume (1754)
[432] Roll (1992), p. 47.
[433] Roll (1992), p. 155.

and lasting influence on businessmen, politicians and generations of future academic economists alike.

Ricardo

The end of the Napoleonic Wars and the economic management of its aftermath proved a stimulating period for the development of economic theory, ushering in the era of what is now known as classical political economy. David Ricardo was by far the most influential economist of this period. His 1817 book *On the Principles of Political Economy and Taxation* developed a theory of economic development that described the dynamics of how the distribution of wealth between landowners, workers and capitalists would change over time (he took the total size of this wealth, the 'national dividend', essentially as a given). Ricardo's *Principles* is recognised as the first attempt at an axiomatic, deductive approach to developing an abstract economic theory, in contrast to the mainly rhetorical and historical style of economic writing that prevailed until then.

Ricardo's theory can be viewed as a reaction to the contemporaneous controversy over the Corn Laws. The Corn Laws were tariffs on the import of grain from overseas, put in place in 1815 to protect domestic farmers and their landowners from overseas competition. To Ricardo, the Corn Laws raised the question of how wealth was distributed between landowners and capitalists. He argued that workers would receive a subsistence wage that would just cover the cost of basic food, irrespective of the cost of that food (and hence irrespective of the size of the food tariff). The Corn Laws therefore would not have a direct impact on worker's quality of life, but they would determine the cost of food, the commensurate level of wages and, consequently, the level of capitalist profits. Meanwhile, the profitability of food production would impact on the rent that landowners could obtain from their land.

The essential conclusions of Ricardian economics were dismally Malthusian (and, indeed, Ricardo and Malthus were close correspondents and later rivals). Malthus' concerns with the sustainability of the significant population growth of the era were given a new economic rigour by Ricardo. Ricardo postulated that agricultural land could be ranked according to its fertility and productivity; farmers would use the most productive land available; as the population increased, more agricultural land would be required to deliver the necessary food supply, and this would mean using increasingly less productive land; the impact of this reduction in the quality of land used would more than offset the improvements in agricultural productivity that could be expected to arise from technological development; the cost of growing food would therefore rise and hence so also would food prices and workers' subsistence wages; in manufacturing, however, technological improvements would reduce the price of goods; business profits would tend to fall as an indirect but inevitable consequence of the increasingly inefficient use of land in less-productive agriculture and the need to provide increased subsistence wages for the larger working population. The economic winner of all this was the landlord, who could take advantage of the demand for his poorer quality land and charge higher rent (which then goes further in its purchasing power for manufactured goods), and hence would obtain a rising share of national income.

Ricardo used his theory to argue for the repeal of the Corn Laws – they merely aggravated the economic problems that his theory showed inevitably arose from a growing population. More generally, Ricardo was a strong advocate of international free trade. The Corn Laws were eventually repealed, but not until 1846, over twenty years after his death.

As an empirically tested deductive scientific theory, Ricardo's theory of economic development failed as fully as any economic theory has ever done – the unfolding reality of the 19th century looked nothing like the predictions of Ricardo's theory. It might be argued this was because some of his key premises – such as that the agricultural productivity increases that could be delivered by science and technology would not be adequate to offset the reduction in the quality of agricultural land that resulted from population growth – proved to be wholly unrepresentative of the emerging reality: 'Ricardo's type of economics, then, consisted of conclusions reached by reasoning from a few broad assumptions that over-simplified the facts…Consequently, most of the conclusions which Ricardo drew about the future have proved to be mistaken.'[434]

The sweeping implications of Ricardo's theory did not require advanced 20th century quantitative techniques for them to be determined as clearly at odds with the unfolding reality of the 19th century. For example, in the decades following the publication of *Principles*, the share of national income paid as rent evidently did not increase as predicted by Ricardo; real wages did significantly increase; and wheat prices fell (even before the repeal of the Corn Laws that eventually arrived in 1846, though, as we will note later, not after).

Nonetheless, Ricardo's work was hugely influential, both for its theoretical content, and, to a lesser but still important degree, for its deductive methodology: as we shall discuss more fully in Chapter 4.2, when its empirical failings became apparent, the theory was defended by economists as a logically self-consistent piece of reasoning whose postulates merely happened to differ from the reality of that particular period in time.

Mil and Schmoller

The next highly notable pillar of classical political economy was erected by John Stuart Mill in the 1830s and 1840s, most notably in his 1848 book *Principles of Political Economy*. Mill's economics can be regarded as a natural development and refinement of Ricardo's. Mill was one of the first philosophers to write extensively and explicitly about methodology in economics. Mill's perspectives on methodology are discussed in some detail in Chapter 4.2. It can be immediately noted that Mill, influence by Comte and an advocate of a positive social science, wrote in support of Ricardo's abstract theory and deductive methods (though Mill argued that universal 'physical' economic laws could only be discovered in the study of the *production* of wealth, but not in its *distribution*[435]; this is especially notable given Ricardo's most famous theory was specifically focused on distribution rather than production).

However, another school of thought, the German historicist school, started to emerge in the 1840s which rejected the idea of economics as an abstract, deductive science. This school reached its greatest influence under Schmoller in the final quarter of the nineteenth century (though it also influenced the American institutionalist school of the 1920s and 1930s founded by Veblen that was similarly sceptical of positive economic methodology and deductive economic theory). Schmoller argued, in an early example of German interpretivism, that only detailed study of concrete historical episodes could lead to economic knowledge.

The Marginal Revolution, the Methodenstreit and the emergence of neoclassical economics

Economics as a discipline of the social sciences underwent a revolutionary change – often referred to as the Marginal Revolution - in the late nineteenth century. It was revolutionary in the sense that it replaced the classical political economy of Smith, Ricardo and Mill with a new and more mathematical approach to abstract, deductive economic theory. Crucial to this shift was the idea of utility theory. More specifically, the concept of marginal utility, a law of diminishing marginal utility and the behavioural postulate of individuals motivated to maximise their utility were the

---

[434] Mitchell in Tugwell (1924), p. 10.
[435] Mill (1848), p. 199.

fundamental ideas that radically altered the nature of economic analysis[436]. This shifted the focus of economic analysis away from a stratified view of society (landowners, workers and capitalists) and a labour cost theory of value to individuals and their consumption preferences and a theory of value based on utility. These ideas were developed independently but contemporaneously by a number of economists in the 1870s, the three most important being W.S. Jevons[437], Leon Walras[438] and Carl Menger[439].

Clearly, the methodological underpinnings of the Marginal Revolution – which involved a commitment to abstract, deductive methods to discover universal laws of a positive economic science - were antithetical to the methodological thoughts of Schmoller and the historical school. This led to a famous methodological dispute, the *Methodenstreit*, between the two schools. This disagreement eventually wore itself out as most reasonable economists started to recognise that the two perspectives could be viewed as complementary rather than mutually exclusive.

The first steps in the 'mathematisation' of economics can arguably be regarded as implicit in Ricardo's embrace of logically deductive abstraction. But the work of the economists of the Marginal Revolution, and most especially Walras, placed increasingly sophisticated mathematical technique at the core of the deductive logic of abstract economic theory. This trend was continued at the start of the twentieth century at Cambridge by Alfred Marshall and, to a greater degree, A.C. Pigou. On the European continent, the work of Vilfredo Pareto[440] in the first decade of the twentieth century was another key strand in the development of mathematical economic theory and neoclassical economics. Marshall, Pigou and Pareto and others such as Irving Fisher in the US collectively refined the theory of the Marginal Revolution into what is now known as neoclassical economics.

The century of historical developments from Ricardo to Pareto can be seen as a long-term program to form economics into a positive scientific discipline with a logico-mathematical content of a scale and sophistication that was (and still is) unique for a social science. Neoclassical economics was deductive, increasingly mathematical, based on methodological individualism (that is, built on behavioural postulates for individuals, and specifically assuming they acted rationally), and showed that free markets tended to quickly obtain equilibria that could be viewed in some specific sense as optimal. Much of neoclassical economics was therefore mainly concerned with 'comparative statics'; that is, the comparison of the equilibria produced under different circumstances.

Keynes

This system received a major challenge from within in 1936 in the form of John Maynard Keynes' *General Theory of Employment, Interest and Money*. Like Ricardo's work, Keynes' *General Theory* can be seen as a reaction to the economic challenges of its age. As noted above, the backdrop to Ricardo's work was the aftermath of the Napoleonic Wars and the implementation of the Corn Laws; in Keynes' case, it was the aftermath of the First World War and the economic depression and mass unemployment that arose in the years that followed which provided the crucial historical context and impetus for his work. The postulates and implications of neoclassical economics did not appear to marry well with the empirical reality of the economic aftermath of the First World War. And that reality was crying out for credible economic guidance.

---

[436] The terminology 'marginal utility' was not, however, introduced until some years later. See Roll (1992), p. 355.
[437] Jevons (1871)
[438] Walras (1874)
[439] Menger (1871)
[440] Pareto (1906)

Keynes was, by this time, a Cambridge economics scholar, schooled particularly in the works of Marshall and Pigou. His *General Theory* was provocative and controversial, and not only to the older generation of more conservative economic thinkers of the neoclassical paradigm. Frank Knight, an acclaimed American contemporary of Keynes, for example, wrote that Keynes 'succeeded in carrying economic thinking well back to the dark ages"[441].

The methodological approach of Keynes' *General Theory* was incongruous to the neoclassical economics that had been established over the fifty years prior to its publication. One of the distinctive features of Keynes' macroeconomics was to move beyond the methodological individualism of neoclassical microeconomics by explicitly considering aggregate phenomena such as total national savings as variables in his theory[442]. Another characteristic of Keynes' approach was his rejection of the idea that markets could be relied upon to efficiently reach static 'optimal' equilibria. His theory was therefore most interested in the dynamic (and potentially long) process by which markets could regain equilibrium (rather than the comparison of alternative static equilibria that characterised neoclassical economic theory).

Keynes sought no quarrel with how neoclassical economic theory's deductive consequences had been logically derived. Rather, his argument was that the basic premises and postulates of neoclassical theory were too far removed from the real economic word. Keynes offered a geometric analogy: neoclassical economics was attempting to apply Euclidean geometry to a non-Euclidean world, and economists had yet to realise that the axiom of parallels must be replaced in order for the subject to achieve empirical success. Coming only a couple of decades after Einstein's (non-Euclidean) theory of relativity, this analogy can be seen as Keynes' attempt to both compliment the neoclassical school (by comparing them with Newton) whilst highlighting that there was a need for a more advanced theory of which the original theory could be considered a special case or approximation. Whilst Newtonian mechanics, however, can be seen to provide a highly accurate approximation to the theory of relativity for the vast majority of practical purposes, Keynes had in mind a theoretical development in macroeconomic theory in which neoclassical economics would be a very restricted special case that he deemed irrelevant to the prevailing macroeconomic environment.

Keynes argued that the postulates of neoclassical economic theory assumed away the possibility of 'involuntary unemployment' (which he defined as where people are unemployed who would be willing to work for less than the then-prevailing real wage). Prior to the *General Theory*, economists' standard explanation for the then-new phenomena of mass unemployment was the 'stickiness' of wages – that is, that workers were reluctant to accept a reduction in money wages, irrespective of what it implied in real terms. Consumer prices experienced a major deflation in the 1920s and early 1930s (in the UK, the Consumer Price Index fell by almost 40% between 1920 and 1934). According to the standard economic explanation, wage 'stickiness' resulted in wages failing to adjust down to their new (real) equilibrium level. A surplus of supply of labour (mass unemployment) arose as an inevitable result of this disequilibrium. From this perspective, a period of inflation could remedy mass unemployment by reducing the real level of wages back down to its equilibrium level.

Keynes agreed that the phenomenon of wage stickiness did exist, but argued it was an inadequate explanation of the growth in mass unemployment since the end of the First World War[443]. Keynes therefore argued that, contrary to the neoclassical theory, reducing nominal money wages via a

[441] Knight (1956), p. 252.
[442] Nagel (1979), p. 544.
[443] Keynes (1936), p. 17.

dose of inflation would not eradicate this new phenomenon of mass unemployment. He argued that the nub of the problem was the neoclassical postulate sometimes referred to as Say's Law, which says that 'the aggregate demand price for all volumes of output as a whole is equal to its aggregate supply price for all volumes of output'. This, he argued, assumes away the possibility of involuntary unemployment arising. Keynes argued that the behaviour of aggregate demand could be more complex than was implied by Say's Law, and this could result in an equilibrium level of employment below full employment.

This led to the perhaps most controversial part of Keynes' theory – that in such times that involuntary unemployment existed, the State could (and indeed should) invest in public works that would raise the prevailing employment equilibrium closer to full employment. Moreover, the increase in employment engendered by this investment would be much greater than the direct employment undertaken by the State – the consumption triggered by their employment would, in turn, create demand that produced further employment. Keynes' theory, and his indicative calibration of it, suggested a 'multiplier' of approximately four would typically apply. The size of the multiplier would be determined by the newly-employed's *marginal propensity to consume* – a law-like description of the consumption and saving tendencies of the population. Keynes noted that this was a *ceteris paribus* argument – if this new investment altered other investment intentions, say, by causing an increase in the interest rate as a result of increased borrowing to fund the investment, the effect would be dampened (and Milton Friedman argued twenty years later that this increase in the interest rate would completely offset the impact on aggregate demand of the increased investment[444]).

The relationship between consumption, investment and employment that Keynes derives from the postulated marginal propensity to consume represents a key element of Keynes' General Theory. Keynes' theory of interest rates represented another key strand of his General Theory. Keynes argued that interest rates were largely determined by liquidity preferences – that is, an individual was induced to exchange cash for a bond by an illiquidity premium. This was distinct from the classical and neoclassical view that the rate of interest was determined by the way the supply of savings and the demand for investment varied as a function of the interest rate. In Keynes' scheme, savings and investment could not be viewed as independently varying quantities for which a clearing rate could be found: instead, both savings and investment were determined by the levels of consumption and income, not the rate of interest. And the rate of interest was not determined by finding the level that matched savings and investment. Rather, savings and investment were identical by definition under the Keynes system, and the rate of interest was determined by the quantity of money and liquidity preferences.

Post-War Economics

Keynes' General Theory had a pervasive global impact on the research agenda of economists for several decades. The macroeconomics of the second half of the twentieth century was dominated by debate between Keynesianism and the competing macroeconomic framework of monetarism, as developed by Milton Friedman and others[445]. Monetarists argue that changes in income (or output) are largely caused by changes in the money supply (via changes in interest rates). Friedman argued that the cause of the Great Depression of the 1930s was not a lack of investment, as was argued by Keynes. Instead, Friedman argued, the cause of the Depression was a great contraction in the money supply. The debate had particular resonance in the 1970s following the oil crisis and recession of

[444] Friedman (1956)

[445] The literature on monetarism is vast, but see, for example, Friedman (1970) for a brief, accessible and informative introduction.

1973 – was Keynesian fiscal stimulus the answer, or, as Friedman would advocate, a monetary stimulus by the central banks the most effective response? This debate remains alive today, both in terms of the theoretical correspondence of the two systems, and their empirical confirmation. But the very existence of this debate highlights a profound change from the pre-Keynesian neoclassical economics of self-correcting markets and economies: some form of government intervention is now regarded as necessary to 'steer' the macro-economy on a path of sustainable growth, the debate is merely about which form of instrument of government intervention is more effective.

Some of the other major themes in post-war economics include the emergence of financial economics (and its especially notable development between Modigliani-Miller in 1958 and Black-Scholes-Merton in 1973); attempts at the development of the 'micro-foundations' of macroeconomics (which sought to explain how 'bottom-up' modelling of individual human behaviour could result in macroeconomic systems; this is now generally viewed as a failed program in the sense that the propositions of the two disciplines of micro and macroeconomics cannot be derived from one another); behavioural economics; and the increasing significance of econometrics as a specialism of economics.

Across several of these topics, the trend of further mathematisation of economic theory is evident. The increase in mathematical content of economics that took place during the quarter-century following the end of the Second World War was described as an 'unprecedented development' in Wasilly Leontief's 1971 Presidential Address to the American Economic Association[446]. It has even been described as a form of Kuhnian revolution (the 'formalist revolution') by some economic writers (though it is a revolution in methodology rather than in the substance of the paradigms of economic theories that are accepted by the economics profession)[447]. The most prominent economists of the second half of the twentieth century such as Paul Samuelson have been instrumental in these developments.

[446] Leontief (1971).
[447] Ward (1972), Chapter 3.

Bibliography

Abbott (1988), *The System of Professions*, University of Chicago Press.

Angrist and Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", *Journal of Economic Perspectives*, Vol. 24, Number 2, 2010.

Ayer (1963), "Two Notes on Probability", *The Concept of a Person*, Macmillan.

Bailey (1862), 'On the Principles on which the Funds of Life Assurance Societies should be Invested', *Journal of the Institute of Actuaries*, Vol. 10, pp. 142-7.

Bayes (1764), "An essay towards solving a problem in the doctrine of chances". *Philosophical Transactions of the Royal Society of London*, 53.

Ben-Haim (2014), 'Order and Indeterminism: An Info-Gap Perspective', in *Error and Uncertainty in Scientific Practice*, Boumans et al (ed), pp. 157-177.

Beck (2014), 'Handling Uncertainty in Environmental Models at the Science-Policy-Society Interfaces', in *Error and Uncertainty in Scientific Practice*, Boumans et al (ed), , pp. 97-137.

Benjamin (1976), 'Maturity Guarantees for Equity-Linked Policies', *20th International Congress of Actuaries*.

Board of Governors of the US Federal Reserve System (2011), *Supervisory Guidance on Model Risk Management*.

Boole (1854), *An Investigation of the Laws of Thought* (2010 Edition), Watchmaker Publishing.

Boyle and Hardy (2003), 'Guaranteed Annuity Options', *Astin Bulletin*, Vol. 33, No. 2, pp. 125-152.

Black, F. and M. Scholes (1973), "The Pricing of Options and Corporate Liabilities", *Journal of Political Economy* 81, No. 3 (May-June 1973), pp. 637-659.

Black, R. et al (2018), 'Model Risk: illuminating the black box', *British Actuarial Journal*, Vol. 23, No. 2.

Blaug (1980), *The methodology of economics*, Cambridge University Press.

Braithwaite (1953), *Scientific Explanation*, Cambridge University Press.

Breiman (2001), "Random Forests", *Machine Learning*, Vol. 24, pp. 123-140.

Breiman, Friedman, Olshen and Stone (1984), *Classification and Regression Trees*, Chapman and Hall.

Bride and Lomax (1994), "Valuation and Corporate Management in a Non-Life Insurance Company", *Journal of the Institute of Actuaries*, Vol. 121, pp. 363-440.

Bridgman (1927), *The Logic of Modern Physics*, Macmillan.

Brown (1963), *Explanation in Social Science*, Routledge.

Browne and Owen (2019), 'Projecting Future Mortality', in *Longevity Risk*, McWilliam et al (ed), Second Edition, pp. 3-42.

Caldwell (1994), *Beyond Positivism*, Routledge.

Carnap (1950), *Logical Foundations of Probability*, Routledge & Kegan Paul.

Carter and Lee (1992), 'Modeling and Forecasting U.S. mortality: Differentials in Life Expectancy by Sex', International Journal of Forecasting, 8, no.3, pp. 393-412.

Cartwright (1983), *How the Laws of Physics Lie*, Oxford University Press.

Cartwright and Montuschi (Editors), *Philosophy of Social Science*, Oxford University Press.

Christ (1975), "Judging the Performance of Econometric Models of the U.S. Economy", *International Economic Review*, 16, pp. 54-74.

Church (1940), "On the Concept of a Random Sequence", Bulletin of the American Mathematical Society, Vol. 46, pp. 130-135.

Cochrane (2005), *Asset Pricing*, Revised Edition, Princeton University Press.

Comte (1875-77), *A System of Positive Polity*, Longmans, Green and Co.

Comte (1910), *A General View of Positivism (translated by J.H. Bridges)*, London.

Cox (1946), "Probability, frequency and reasonable expectation", *American Journal of Physics*, Vol. 14, p. 1-13.

Cox (1961), *The Algebra of Probable Inference*, John Hopkins University Press.

CRO Forum (2006), 'A market cost of capital approach to market value margins'.

Dalkey & Helmer (1963), 'An Experimental Application of the Delphi Method to the use of experts', *Management Science*, Vol. 9, p. 458-467.

Daston (1988), *Classical Probability in the Enlightenment*, Princeton University Press.

De Finetti (1970), *Theory of Probability*, Wiley. (page references refer to 2017 edition)

De Morgan (1847), *Formal Logic*, Taylor and Walton.

Dirac (1958), *The Principles of Quantum Mechanics*, 4th Edition, Oxford University Press.

Duhem (1904-5), *The Aim and Structure of Physical Theory*, Atheneum.

Dunn (1961), 'Multiple Comparisons Among Means', *Journal of the American Statistical Association*, Vol. 56, p. 52-64.

Efron and Hastie (2016), *Computer Age Statistical Inference*, Cambridge University Press.

Einstein (1933), *On the Method of Theoretical Physics*, Oxford.

Exley, Mehta and Smith (1997), 'The Financial Theory of Defined Benefit Pension Schemes', *British Actuarial Journal*, Vol. 3, No. 4, pp. 835-966.

Fama and French (1988), 'Permanent and Temporary Components of Stock Prices', *Journal of Political Economy*, Vol. 96, pp. 246-273.

Feigl (1953), "Notes on Causality", in Feigl and Brodbeck, *Readings in the Philosophy of Science*, Appleton-Century-Crofts.

Feyerabend (2010), *Against Method*, Fourth Edition, Verso.

Fisher, I. (1911), *The Purchasing Power of Money*, Macmillan.

Fisher, R. (1922), 'On the mathematical foundations of theoretical statistics', *Philosophical Transactions of the Royal Society*, A, pp. 309-368.

Fisher, R. (1935), *The Design of Experiments*, Oliver and Boyd.

Fisher, R. (1956), *Statistical Methods and Statistical Inference*, Oliver and Boyd.

Ford et al (1980), "Report of the Maturity Guarantee Working Party", *Journal of the Institute of Actuaries*, Vol. 107, pp. 103-231.

Fraasen (1980), *The Scientific Image*, Clarendon Press.

Frege (1879), *Begriffsschrift: Eine der arithmetischeen nachgebildete Formelsprache de reinen Denkens*. Halle.

Friedman (1953), The Methodology of Positive Economics, *Essays in Positive Economics*, University of Chicago.

Friedman (1956), "The Quantity Theory of Money: A Restatement", *Studies in the Quantity Theory of Money*, University of Chicago Press.

Friedman (1976), *Price Theory*, Aldine Transaction.

Friedman (1970), "The Counter-Revolution in Monetary Theory", *Institute of Economic Affairs*.

Frisch (1933), "Editorial", *Econometrica*, Vol. 1, pp. 1-4.

Gao, Meng and Wuthrich (2018), "Claims Frequency Modeling Using Telematics Car Driving Data", *SSRN*.

Geoghegan et al, 'Report on the Wilkie Investment Model', *Journal of the Institute of Actuaries*, Vol. 199, pp. 173-228.

Gibson (1960), *The Logic of Social Enquiry*, Routledge.

Giere (1988), *Explaining Science: A Cognitive Approach*, University of Chicago Press.

Gilley (1972), 'The Dissolution of a Pension Fund', *Journal of the Institute of Actuaries*, Vol. 98, No. 3, pp. 179-232.

Gillies (1993), *Philosophy of Science in the Twentieth Century*, Blackwell.

Gillies (2000), *Philosophical Theories of Probability*, Routledge.

Gompertz (1825), "On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Model of Determining the Value of Life Contingencies", *Philosophical Transactions of the Royal Society of London*, Vol. 115, pp. 513-583.

Gompertz (1871), 'On one Uniform Law of Mortality from Birth to extreme Old Age, and on the Law of Sickness', *Journal of the Institute of Actuaries*, Vol. 16, No. 5, pp. 329-344.

Good (1956), "Which Comes First, Probability or Statistics?", *Journal of the Institute of Actuaries*, Vol. 82, p. 249.

Grandy (1973), *Theories and Observation in Science*, Ridgeview.

Granger (1969), “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”, *Econometrica*, 37, pp. 424-38.

Graunt (1662), *Natural and Political Observations Mentioned in a Follow Index, and Made Upon the Bills of Mortality*.

Haavelmo (1947), “Methods of Measuring the Marginal Propensity to Consume”, *Journal of the American Statistical Association*, 42, pp. 105-22.

Hacking (1965), *The Logic of Statistical Inference*, Cambridge University Press.

Hacking (1975), *The Emergence of Probability*, Cambridge University Press.

Hainaut (2017), “A Neural-Network Analyzer for Mortality Forecast”, *Astin Bulletin*, Vol. 48, No. 2, pp. 481-508.

Haldane and Madouros (2012), “The Dog and the Frisbee”, *Proceedings – Economic Policy Symposium – Jackson Hole*, pp. 109-159.

Hardy (2006), *An Introduction to Risk Measures for Actuarial Applications*, Casualty Actuarial Society and Society of Actuaries.

Harrod (1938), “The Scope and Method of Economics”, Economic Journal, Vol. 48, pp. 383-412.

Hausman (1992), *The inexact and separate science of economics*, Cambridge University Press.

Hausman (2008), *The Philosophy of Economics: An Anthology*, Third Edition, Cambridge University Press.

Hayek (1975), “Full Employment at Any Price?”, *The Institute of Economic Affairs*, Occasional Paper 45.

Hempel (1959), *The Function of General Laws in History*, in *Theories of History*, ed. by Gardner, Collier-Macmillan.

Hempel (1962), “Deductive-Nomological vs. Statistical Explanation”, *Minnesota Studies in the Philosophy of Science*, 3, pp. 98-169.

Hempel and Oppenheim (1948), “Studies in the Logic of Explanation”, Philosophy of Science, Vol. 25, No. 2, pp. 135-175.

Hesslow (1976), “Discussion: Two Notes on the Probabilistic Approach to Causality”, *Philosophy of Science*, Vol. 43, pp. 290-292.

Hicks (1969), *A Theory of Economic History*, Oxford University Press.

Hicks (1979), *Causality in Economics*, Blackwell.

HM Treasury (2013), *Review of quality assurance of Government analytical models: final report*.

Hooper, Mishkin and Sufi (2019), “Prospects for Inflation in a High Pressure Economy: Is the Phillips Curve Dead or is It Just Hibernating?”, *2019 US Monetary Policy Forum*.

Hoover (1994), “The Lucas Critique and the Nature of Econometric Inference”, *Journal of Economic Methodology*, vol. 1, pp. 65-80.

Hoover (2001), *Causality in Macroeconomics*, Cambridge University Press.

Hotelling (1927), "Differential Equations Subject to Error, and Population Estimates', *Journal of the American Statistical Association*, Vol. 22, pp. 283-314.

Howson and Urbach (1993), *Scientific Reasoning: The Bayesian Approach, Second Edition*, Open Court Publishing.

Huber (1996), "A Study of the fundamentals of actuarial economic models", Unpublished Doctoral thesis, *City University London*.

Hume (1740), *A Treatise of Human Nature*, John Noon.

Hume (1748), *An Enquiry Concerning Human Understanding*, Oxford World's Classics.

Hume (1754), *Essays: Moral, Political and Literary*, Liberty Classics.

Hutchison (1938), *The Significance and Basic Postulates of Economic Theory*, Augustus M. Kelley.

Institute of Actuaries (1996), *Manual of Actuarial Practice*, Supplement 3.

Institute and Faculty of Actuaries (2016), *Strategy 2016*.

Institute and Faculty of Actuaries (2017), *Longevity Bulletin*.

Institute of Actuaries and Faculty of Actuaries (1979), *a(90) tables for annuitants*.

International Association of Insurance Supervisors (2018), *Risk-based Global Insurance Capital Standard Version 2.0*, *Public Consultation Document*.

Ioannidis (2005), "Why Most Published Research Findings Are False", *PLoS Medicine*, Vol. 2, No. 8, pp. 696 – 701.

Jaynes (2004), *Probability Theory: The Logic of Science*, Cambridge University Press.

Jeffreys (1937), *Scientific Inference*, Second Edition, Cambridge University Press.

Jeffreys (1961), *Theory of Probability,* Third Edition, Oxford. (Page references refer to 1998 Oxford Classics series).

Jevons (1871), *The Theory of Political Economy*, Macmillan.

Johnson (1924), *Logic, Part III*, Cambridge University Press.

Jones (1965), "Bayesian Statistics", *Transactions of the Society of Actuaries*, XVII.

Jones and Kimeldorf (1967), "Bayesian Graduation", *Transactions of the Society of Actuaries*, XIX.

Karmack (1983), *Economics & the Real World*, Blackwell.

Keynes, J.M. (1921), *A Treatise on Probability*, Macmillan & Co. (Page references refer to 2007 Watchmaker Publishing edition).

Keynes, J.M. (1936), *The General Theory of Employment, Interest and Money*, Wordsworth Editions.

Keynes, J.M. (1973), *The Collected Writings of John Maynard Keynes, Vol. XIV*, Cambridge University Press.

Keynes, J.N. (1891), *The Scope and Method of Political Economy*, Macmillan.

King (2016), *The End of Alchemy*, Little, Brown.

Knight (1921), *Risk, Uncertainty and Profit*. Hart, Schaffner & Marx.

Knight (1935), Value and Price, from *The Ethics of Competition and Other Essays*, Harper and Brothers.

Knight (1956), *On the History and Method of Economics*, University of Chicago Press.

Knuuttila (2009), 'Representation, Idealization and Fiction in Economics', in *Fictions in Science*, ed. Suarez, Routledge.

Kolmogorov (1950), *Foundations of the Theory of Probability*, Chelsea Publishing Company.

Koopmans (1957), *Three Essays on the State of Economic Science*, McGraw-Hill.

Krugman (2009), "How Did Economists Get It So Wrong?", New York Times, 2nd September.

Kuhn (1996), *The Structure of Scientific Revolutions*, Third Edition, The University of Chicago Press.

Kyburg and Smokler (1980), *Studies in Subjective Probability, Second Edition,* Krieger Publishing.

Ladyman (2002), *Understanding Philosophy of Science*, Routledge.

Lakatos and Musgrave (1970), *Criticism and the Growth of Knowledge*, Cambridge University Press.

Laplace (1820), *Theorie Analytique des Probabilites*, Courcier Imprimeur.

Laudan (1981), 'A Confrontation of convergent realism', *Philosophy of Science*, Vol. 48, pp. 19-48.

Lawson (1999), 'What Has Realism Got to Do with It?', *Economics and Philosophy*, Vol. 15, pp. 269-82.

Lawson and Pesaran (1985), *Keynes' Economics: Methodological Issues*, Croom Helm.

Leontief (1971), 'Theoretical Assumptions and Nonobserved Facts', *American Economic Review*, 61(1), pp. 1-7.

Lichtenstein and Slovic (1971), 'Reversals of Preference Between Bids and Choices in Gambling Decisions', *Journal of Experimental Psychology*, Vol. 89, pp. 46-55.

Linnebo (2017), *Philosophy of Mathematics*, Princeton University Press.

Lucas (1976), 'Econometric policy evaluations: a critique', in Bruner and Meltzer (eds) *The Philips Curve and Labor Market*, North Holland.

Machlup (1978), *Methodology of Economics and Other Social Sciences*, Academic Press.

Maki (1994), 'Reorienting the Assumptions Issue', in *New Directions in Economic Methodology*, Routledge.

Manly (1911), 'On Staff Pension Funds: The Progress of the Accumulation of the Funds…', *Journal of the Institute of Actuaries*, Vol. 45, No. 2, pp. 149-231.

Mayer-Schonberger and Cukier (2017), *Big Data*, John Murray.

McLeish (1983), 'A Financial Framework for Pension Fund Finance', *Transactions of the Faculty of Actuaries*, Vol. 38, pp. 267-314.

McLeish and Stewart (1987), ‘Objectives and Methods of Funding Defined Benefit Pension Schemes’, *Journal of the Institute of Actuaries*, Vol. 114, No. 2, pp. 155-225.

Menger (1871), *Grundsatze der Volkswirtschaftslehre*, Vienna.

Merton, R.C. (1973), “Theory of Rational Option Pricing”, *Bell Journal of Economics and Management* 4, No. 1 (Spring 1973), pp. 141-183.

Merton (1974), “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”, *Journal of Finance*, Vol. 29, pp. 449-470.

Mill (1836), “On the Definition of Political Economy; and on the Method of Investigation Proper to It”. In *Essays on Some Unsettled Questions of Political Economy*. Parker. Page references refer to edition published in Hausman (2008).

Mill (1848), *Principles of Political Economy*.

Mill (1879), *A System of Logic*, London.

Mill (1959), *Elucidations of the Science of History*, in *Theories of History*, ed. by Gardner, Collier-Macmillan.

Morgan (1992), *The History of Econometric Ideas*, Cambridge University Press.

Morgan and Morrison (1999), *Models as Mediators*, Cambridge University Press.

Morjaria et al (2016), ‘Model Risk: Daring to open up the black box’, *British Actuarial Journal*, Vol. 21, No. 2.

Nagel (1939), *Principles of the Theory of Probability*, International Encyclopaedia of Unified Sciences.

Nagel (1959), *Some Issues in the Logic of Historical Analysis*, in *Theories of History*, ed. by Gardner, Collier-Macmillan.

Nagel (1979), *The Structure of Science*, Hackett.

Neyman (1937), ‘Outline of a theory of statistical estimation based on the classical theory of probability’, *Philosophical Transactions of the Royal Society*, A, pp. 333-80.

Neyman and Pearson (1933), ‘On the problem of the most efficient tests of statistical hypotheses’, *Philosophical Transactions of the Royal Society*, A, pp. 289-337.

Noll, Salzmann and Wuthrich (2018), “Case Study: French Motor Third-Party Liability Claims”, *SSRN*.

Nowell et al (1999), *Continuous Mortality Investigation, Report 17*, Institute of Actuaries and Faculty of Actuaries.

Nye (1972), *Molecular Reality*, American Elsevier.

Ogborn (1962), *Equitable Assurances*, George Allen and Unwin.

Oliver (2019), ‘Determinants of Changes in Life Expectancy’, in *Longevity Risk*, McWilliam et al (ed), Second Edition, pp. 3-42.

Ouchin (2004), “A Literature Review on the Use of Expert Opinion in Probabilistic Risk Analysis”, *World Bank Policy Research Working Paper 3201*.

Pap (1946), *The* A Priori *in Physical Theory*, King’s Crown Press.

Pareto (1906), *Manual of Political Economy*, Oxford University Press.

Pearson (1911), *The Grammar of Science*, 3rd Edition. (Page references refer to Cosimo (2007) print).

Pemberton (1999), "The Methodology of Actuarial Science", *British Actuarial Journal*, Vol. 5, pp. 115-196.

Pension Protection Fund (2018), *The Purple Book 2018*.

Petty (1899), *The Economic Writings of Sir William Petty*, Cambridge University Press.

Phillips (1958), "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957), *Economica*, November 1958.

Planck (1959), *The New Science*, New York Meridian.

Poincare (1902), *Science and Hypothesis*, English translation, Dover, 1952.

Popper (1957), *The Poverty of Historicism*, Routledge.

Popper (1959), *Prediction and Prophecy in the Social Sciences*, in *Theories of History*, ed. by Gardner, Collier-Macmillan.

Popper (1959), 'The Propensity Interpretation of Probability', *British Journal for the Philosophy of Science*, Vol. 10, pp. 25-42.

Popper (1959a), *The Logic of Scientific Discovery*. Hutchison.

Price (1772), *Observations on Reversionary Payments...,* T. Cadfll.

Quine (1953), *From a Logical Point of View*. Harvard University Press.

Redington (1986), *A Ramble Through the Actuarial Countryside*, Staple Inn.

Reiss (2013), *Philosophy of Economics*, Routledge.

Ricardo (1817), *On the Principles of Political Economy and Taxation*.

Richman (2018), "AI in Actuarial Science", *SSRN*.

Robbins (1935), *An Essay on the Nature and Significance of Economic Science*, 2nd edition, Macmillan. Page references refer to edition published in Hausman (2008).

Roberts (1996), *The Logic of Historical Explanation*, Pennsylvania State University Press.

Roll (1988), 'R^2', *Journal of Finance*, Vol. 43, No. 3, pp. 541-566.

Roll (1992), *A History of Economic Thought*, Fifth Edition, Faber and Faber.

Russell (1912), *The Problems of Philosophy*, Williams and Norgate.

Russell (1912-13), 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, Vol. 13.

Russell (1918), *Mysticism and Logic*, Allan and Unwin.

Ryder (1976), "Subjectivism – A Reply in Defence of Classical Actuarial Methods", *Journal of Institute of Actuaries*, Vol. 103, No. 1, pp. 59-112.

Ryder (1981), 'Consequences of a Simple Extension of the Dutch Book Argument', *British Journal for the Philosophy of Science*, Vol. 32, p. 164-7.

Salmon (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.

Savage (1972), *The Foundations of Statistics, 2nd edition,* Dover Publications.

Schmueli (2010), "To Explain or to Predict", *Statistical Science*, Vol. 25, No. 3, pp. 289-310.

Schultz (1971), "Has the Phillips Curve Shifted? Some Additional Evidence", *Brookings Papers on Economic Activity*, 2, 1971.

Schumpeter (1949), 'The Historical Approach to the Analysis of Business Cycles', *Universities-National Bureau Conference on Business Cycle Research*, November 1949.

Shiller (1981), 'Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?', *American Economic Review*, Vol. 71, No. 3, pp. 421-436.

Siegel (2016), *Predictive Analytics*, Wiley.

Simon (1953), "Causal Ordering and Identifiability", in Hood and Koopman, *Studies in Econometric Method*, Wiley & Sons.

Simon (1956), "Rational Choice and the Structure of the Environment", Psychological Review, 63(2), pp. 129-138.

Smith, A. (1776), *An Inquiry into the Nature and the Causes of the Wealth of Nations,* Wordsworth Editions.

Smith, V. (1994), 'Economics in the Laboratory', *Journal of Economic Perspectives*, vol. 8, pp. 113-31.

Sombart (1902), *Der modern Kapitalism*, Munich.

Soros (2003), *The Alchemy of Finance*, Wiley.

Suarez (2009), *Fictions in Science*, Routledge.

Sugden (2000), 'Credible Worlds: The Status of Theoretical Models in Economics', *Journal of Economic Methodology*, Vol. 7, pp. 1-31.

Susskind and Susskind (2017), *The Future of the Professions*, Oxford University Press.

Swinburne (1974), *The Justification of Induction*, Oxford University Press.

Taylor (1969), *The Trouble-Makers: Dissent over Foreign Policy, 1729-1939*. London.

Thornton and Wilson (1992), 'A realistic approach to pension funding', *Journal of the Institute of Actuaries*, Vol. 199, pp. 299-312.

Tinbergen (1939), *Statistical Testing of Business Cycle Theories, Vol. I and Vol. II*. League of Nations.

Tugwell (1924), *The Trend of Economics*, Alfred Knopf.

Turnbull (2017), *A History of British Actuarial Thought*, Palgrave Macmillan.

Turnbull (2018), 'Some Notes on Approaches to Regulatory Capital Assessment for Insurance Firms', *British Actuarial Journal*, Volume 23.

Walras (1874), *Elements d'economie pure; on theorie mathematique de la richesse sociale*. Lausanne.

Van Fraassen (1980), *The Scientific Image*, Oxford University Press.

Vasicek, O. (1977), "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics* 5, pp. 177-188.

Venn (1888), *Logic of Chance*, Third Edition, Macmillan and Co. (Page references use 2015 edition by Forgottenbooks.com)

Venn (1889), *The Principles of Empirical or Inductive Logic*, Macmillan and Co.

Von Neumann and Morgenstern (1947), *Theory of Games and Economic Behaviour*, 2nd Edition, Princeton.

Von Mises (1928), *Probability, Statistics and Truth*, Allen and Unwin.

Von Mises (1949), *Human Action: A Treatise on Economics*, Yale University Press.

Wald (1939), 'Contributions to the Theory of Statistical Estimation and Testing Hypotheses', *Annals of Mathematical Statistics*, Vol. 10, No. 4, pp. 299-326.

Ward (1972), *What's Wrong with Economics?*, Macmillan.

Weber (1927), *General Economic History*, Greenberg.

Weber (1930), *The Protestant Ethic and the Spirit of Capitalism*, Allen and Unwin.

Weber (1949), *The Methodology of the Social Sciences*, Macmillan.

Weiss et al (2019), 'Ageing Populations and Changing Demographics', in *Longevity Risk*, McWilliam et al (ed), Second Edition, pp. 3-42.

Wilkie (1984), 'A Stochastic Investment Model for Actuarial Use', *Transactions of the Faculty of Actuaries*, Vol. 39, pp. 341-403.

Wilkie (1995), 'More on a Stochastic Asset Model for Actuarial Use', British Actuarial Journal, Vol. 1, pp. 777-964.

Wilkie et al (1990), *Continuous Mortality Investigation, Report 10*, Institute of Actuaries and Faculty of Actuaries.

Wuthrich (2017), "Covariate Selection from Telematics Car Driving Data", *European Actuarial Journal*, Vol. 7, pp. 89-108.

Young (1880), "Can a Law of Mortality be Represented in a Mathematical Form", *Journal of the Institute of Actuaries*, 22, pp. 139-40.

Young (1897), "The Nature and History of Actuarial Work as exemplifying the Mode of Development and the Methods of Science", *Journal of the Institute of Actuaries*, 33, pp. 97-131.

Zellner (1980), *Bayesian Analysis in Econometrics and Statistics*, North-Holland.

Zellner (1984), *Basic Issues in Econometrics*, University of Chicago Press.