To be read to the Royal Statistical Society at the RSS Annual Conference, at 9am on 9th September 2026, Bournemouth International Centre

Bayesian Statistics in Scientific Induction
A Decision Analysis Perspective

Simon French[1]
Alliance Manchester Business School
University of Manchester
Booth Street West
Manchester, M15 6PB, UK

September 2025

Abstract

Based on over fifty years' experience of applying Bayesian methodology in statistical, risk and decision analyses, I reflect on its power in supporting sound scientific induction. I argue that Bayesian methods are more suited to guiding inferences than other methods. In the past, Bayesians justified their approach through its rigorous axiomatic base and its intuitive good sense in modelling rational inference. Nowadays, the easily comprehended interface offered by belief nets, the computational power of MCMC to conduct complex analyses, and the general fit of Bayesian methodology with machine learning are more likely to be offered as justification. Here I offer a further perspective, focusing on the support Bayesian analysis offers scientific induction. Using Shafer's interpretation of prescriptive analysis as 'argument by analogy', I argue that Bayesian analysis is better suited to nudging scientists towards rational inference in confirmatory studies and reducing the risk that they may be misled by subconscious biases. However, I also recognise that non-Bayesian approaches can be very useful in exploratory analyses. Moreover, I recognise that subjective probabilities cannot model all the uncertainties that must be addressed in sound scientific induction. The unmodelled ones, I suggest, need much more recognition and attention than apparent in current practice. Throughout, I recognise that the process of scientific induction is – or ideally would be – driven by a series of interactions between scientists and statisticians. Thus the paper presents my personal perspective on how a subjective statistician should engage with scientists to support their scientific induction.

Keywords: Bayesian analysis; Prescriptive Analysis; Scientific Induction; System 1 and System 2 Thinking; Uncertainty.

1 Introduction

I became a Bayesian during 1971 in the final year of a mathematics degree. I had taken an introductory statistics course in my second year, but neither enjoyed nor really understood it. However, in the following year a course based on DeGroot's (1970) *Optimal Statistical Decisions* captured my interest. I adopted its philosophy and have labelled myself as a Bayesian ever since. My doctoral studies used Bayesian three stage models and variations on the Kalman filter to analyse crystallographic data on

[1] Email for correspondence: simon.french.50@gmail.com. ORCID #: 0000-0002-2013-8571

proteins (French, 1978; French & Wilson, 1978; Oatley & French, 1982). Put simply, Bayesian methods led to far more interpretable and useful electron density maps than previous 'frequentist adhockeries', if I may use that pejorative term. Although I continued to follow and write on Bayesian statistics, my subsequent career moved more into risk and decision analysis, where I developed and applied Bayesian methods. Then in 2011 I semi-retired and joined the Statistics Department at Warwick University part-time, leading their consultancy work until I retired fully in 2020. This reawakened my interests in statistical foundations and scientific induction and provided considerable experience of applied projects. Moreover, with the newly found freedom to reflect and think brought by partial, then full retirement, I have been re-exploring and re-evaluating Bayesian ideas based on a half-century of their use (French, 2013, 2015, 2021, 2022, 2023, 2024; French & Argyris, 2018). In particular, in French (2024) I explore differences in application between Bayesian ideas in statistical, risk and decision analysis. Here, I am concerned solely with Bayesian inference in scientific induction.

I am still a Bayesian, there is no doubt of that; but now my perspective is softer. Though once I engaged actively in the almost gladiatorial discussions between Bayesians and non-Bayesians, I now have a more integrative view. I accept the need for 'horses for courses', in the vernacular, but I do define the courses rather carefully!

Experience has taught me that there are substantial differences in detail, though not philosophy, in the application of Bayesian ideas in statistical, risk and decision analysis (French, 2024). In this paper, I want to discuss Bayesian statistical methodology from the perspective of the theory and practice of decision analysis. Savage (1972), after all, laid the foundations for Bayesian statistics on the basis of an axiomatic development of rational decision-making. I do not wish to get into any technical detail, reflecting instead on how the analysis supports problem-solving, in this case learning from data and building consensual scientific knowledge. I shall make a distinction between *inference* from a specific data set and the broader process of *scientific induction*, in which putative hypotheses describing possible workings of the external world are created, experiments or surveys designed that may be informative on which of these might be true in some sense, data analysed and inferences made, then peer reviewed before becoming part of consensual scientific knowledge. You may gather from my use of 'consensual' and my stated adherence to Bayesian methodology that I am a subjectivist. I am uncomfortable with any concept of objective scientific knowledge, believing that scientific knowledge grows as consensus forms among individual scientists. This paper may perhaps be best described as a personal perspective on how a subjective statistician should engage with scientists to support their work throughout the entire process of scientific induction.

I have several specific objectives:

i. to suggest that the statistician should not just be a backroom analyst churning through data, but should adopt a more active role, facilitating the entire process of scientific induction;
ii. to increase the recognition of the role of statistical analysis in countering potential subconscious biases and nudge those involved to more rational inferences and sound scientific induction;
iii. to argue that all statisticians, not just Bayesians, ignore some of the uncertainties that need addressing in designing, managing and reporting a research project, which may have contributed to some of the replication issues that are becoming apparent in research;
iv. to encourage the statistical profession to promote much clearer reporting of uncertainty so that the users of the results of scientific investigations are fully aware of all the remaining sources of uncertainties.

In the next section, I discuss modelling and analysis and where uncertainties may enter – or be created by – the process. In Section 3, I summarise the principles of Bayesian analysis as they were commonly accepted in the 1970s and 1980s, contrasting them with an admittedly uncommon view of other statistical schools at that time. The remainder of the paper will discuss how Bayesian techniques can be employed prescriptively to support scientists seeking to collect and analyse data to increase their understanding of the world. In Section 5, I turn to a discussion of uncertainty, the probability modelling of which is inevitably part of statistical inference. But there are many forms of uncertainty that are *not* amenable to probability modelling or even quantitative sensitivity and robustness analyses. These are often ignored in much statistical practice, but addressing them in the process of scientific induction is vital if our knowledge of the world is to develop in a sound and balanced manner. The Cynefin Model, introduced in Section 6, offers an important perspective on uncertainty relating its quality to our knowledge of cause and effect. Recognising how Cynefin would categorise the context of a scientific investigation provides an informative setting for discussing exploratory and confirmatory studies and the roles of Bayesian and non-Bayesian methods in inference. In Section 7, I turn to problem formulation or *soft elicitation*, a topic that is seldom discussed in depth in the statistical literature, though its effective conduct is essential to a sound inference. How does one work with the scientists to determine the full context of their investigation, their aims and objectives, etc. so that the statistical modelling and analysis is as appropriate as possible to answer their questions of concern?

These early sections introduce the concepts, ideas and perspective needed for the main discussion of this paper: how do Bayesian and other analyses fit into and facilitate the process of scientific induction? In Section 8, I outline a process-oriented view of a scientific investigation, suggesting how a Bayesian analysis should be developed and explored, recognising that making any inference requires a *judgement* that the analysis is requisite, i.e. good enough to draw a conclusion that may be offered for reporting and peer review. I also suggest that the statistician, being independent of the research itself, is uniquely positioned within a research project to facilitate discussions and deliberations within the project. In Section 9, I argue that for the peer-review process to be sound, the reporting needs to reflect more on the uncertainties that are *not* modelled within the statistical analysis and, of course, to do that one needs to pay attention to how these are communicated in the report. Section 10 offers further discussion and Section 11 summarises the conclusions.

2 Modelling and Analysis

George Box is famous for his aphorism that "all models are wrong, but some are useful," which he used in many discussions with slight variations in wording. It is a remark that has reached many communities far from Statistics. Yet despite his warning, many believe in their models, living within them almost unquestioningly. The modern terminology coming from the computing fraternity in referring to complex models of systems as *digital twins* and the literature surrounding those is an indication of the growing belief in the near perfection of some models (Jones et al., 2020). So before turning to the substance of my arguments, I want to reflect quickly on some issues relating to models and how we use them, suggesting that if we are to find them useful, we need to recognise them as *approximations* to some world. Thus we need to be aware that they may not, indeed *do not* capture *all* elements of that world.

Models describe things: aspects of possible worlds, real or imaginary. They describe entities and behaviours between these. Models encode knowledge and theories, personal and scientific; and they can be used in many ways from describing some aspect and behaviour of the world as it is to some

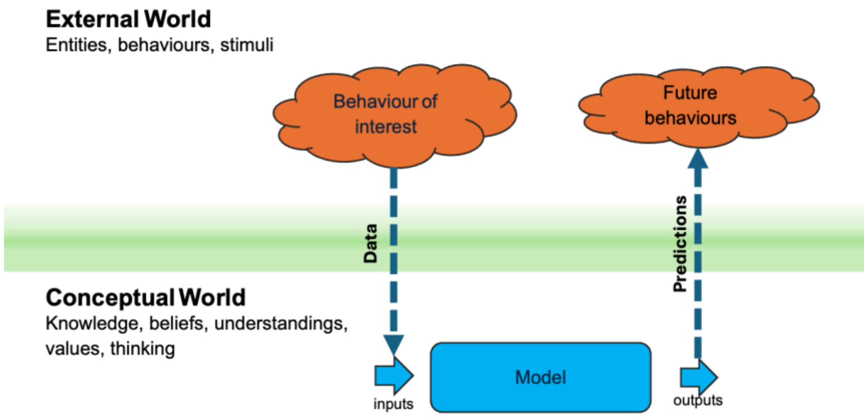

Figure 1: The relationship between the real external world and descriptive models

possible future or entirely imaginary, idealised state. Given our focus on Bayesian statistical methodology, I shall consider descriptive and normative models. In simple terms, d*escriptive* models describe the world and people as they *are* and *do behave*; n*ormative* models describe the world and, particularly, people as they *should behave*. In a statistical analysis we seek to compare a putative descriptive model with data using a normative model of rational inference.

We shall need to distinguish the real, complex, external world where the behaviours, stimuli and things that affect our lives exist from our internal conceptual world in which our knowledge, beliefs, understanding, values exist and our thinking takes place (Figure 1). Our conceptual world holds our perception of entities in the external world and our understanding of their interactions, i.e. causes and effects. Philosophers have spent centuries discussing the relationships between the external and conceptual worlds without reaching any strong consensus, save that they are separate. So here I will be pragmatic and simply assume that in aligning the two worlds, we make judgements. Since quantification can only exist in the conceptual world[2], these judgements cannot be quantified. We need to recognise this if we are to understand the uncertainty that remains after any statistical analysis.

Within our conceptual world, we build models to explain and embody our knowledge of external behaviours, their causes and effects. Models transform inputs -- roughly, 'causes' – into outputs – roughly, 'effects'. A model is not a perfect representation of the external world, entities and their behaviour; nor are data perfect representations of the stimuli from that behaviour. The predictions are an interpretation of the model outputs and again we cannot be sure that they align perfectly with actual future behaviours in the world. Thus there are inevitably unmodelled errors and uncertainties in the processes represented by the dashed lines in Figure 1. These errors will reduce as our knowledge and sophistication of our modelling improves, but conceptually they can never disappear without completely identifying the model and the external world as the same thing. Thus, although modelling and statistical analysis are processes which help us reduce many of our uncertainties about what is happening in the external world, the processes of increasingly sophisticated modelling and abstracting data from the external world as we learn about the world and those of interpreting analyses introduce hopefully smaller, unquantified and conceptually unquantifiable modelling

[2] Some suggest numbers and hence quantification exists in the external world, but while you can have, say, four cows or six geese, you cannot find 'four' or 'six' or any numbers as entities in their own right. They are conceptual descriptions in the conceptual world.

uncertainties, which we need to acknowledge. So to repeat, models are approximations and we ignore the difference between them and the real world at our peril; and this remark applies to all analyses, Bayesian or non-Bayesian.

Normative models do not correspond with the real external world, but an imagined idealised one in which some behaviours of individuals are always consistent with some conception of rational behaviour. In the case of Bayesian inference and decision those assumptions underpinning such consistency are well studied, as indeed they are in a few other paradigms (DeGroot, 1970; Fine, 1973; French, 2022; French & Rios Insua, 2000; Roberts, 1979; Savage, 1972). Normative modelling, in the case of statistical inference, allows us to investigate what an idealised statistician would infer from a set of data in the context of some model world. We shall expand on this in our discussion of prescriptive analysis in Section 4.

3 Bayesian and non-Bayesian Positions in the 20th Century

During the early years of my career, discussions between Bayesians and non-Bayesians were, to say the least, intense, full of polemics. Barnett (1999) provides a balanced account of both sides of the debate; and I do not wish to rehearse those arguments here. Nowadays, Bayesian methodology is well accepted in the professional statistical community, though perhaps less so in some research communities where statistical textbooks for specific disciplines and journal editorial policies can sometimes seem to be stuck in a time-warp. By and large though, computational advances such as Markov Chain Monte Carlo (MCMC) and graphical interfaces have made Bayesian methods more tractable, and the results have won many converts.

Ramsey (1926), De Finetti (1937; 1974, 1975) and Savage (1972) laid the axiomatic foundations of the Bayesian paradigm, discussing rationality and providing arguments to justify the use of Bayes' Theorem and Subjective Expected Utility (SEU) in inference and decision. By the late 20th century, The Bayesian approach was encapsulated by Denis Lindley (1992) as:

> *"It is sensible to begin with a statement of the Bayesian position so that it is clear what is under discussion. That position has three features:*
>
> *(a) All quantities whose value is not known to you ought to have their uncertainties described by numbers that obey the rules of the probability calculus. Thus the correct measure of the uncertainty of a hypothesis is the probability of that hypothesis, not a significance level. The uncertainty of a defendant's guilt should similarly be a probability of guilt.*
> *(b) All consequences should have their worth to you described by numbers on a utility scale; a scale itself based upon probability. The relationship between utility and probability is essential for the third feature.*
> *(c) Decisions, or acts, should be based on the maximization of expected utility; the expectation being in accord with the probabilities described in (a)."*

His point (a) emphasises that in an analysis *all* uncertainties should be encoded in probability. Arguing against that as a possibility is one of the key purposes of this paper, even though I should dearly like it to be true. In the case of (b), I would agree wholeheartedly with the use of utility in all analyses and, indeed, utility models are used commonly in risk and decision analyses. That is not the case, currently, in statistical practice. In many cases, Bayesians argue that for inference all that is needed is to report the posterior distribution, and when they do use utility – or rather its negative, a loss function – they usually adopt a conventional representation of loss, e.g. squared error or logarithmic, rather than elicit one tailored to the scientists' actual concerns about error, using methods such as those discussed in Keeney and Raiffa (1993).

As an author of a text on *statistical decision theory*, I have a slight leaning towards using appropriate loss functions (French & Rios Insua, 2000). I recognise, however, that many better Bayesians than I take a purely inferential approach (Box & Taio, 1973; Christensen et al., 2011; Gliner et al., 2009; Spiegelhalter, 2019). As a long-practising decision theorist and analyst, I have an instinct to think about how the new knowledge will be used. So we should acquire it in ways that mean it is suitable for use: e.g., errors of over or under estimation may have different importance that should be recognised by using a suitable loss function. In practice, though, there is little difference between the approaches. To evaluate the expected loss, the first step is to calculate the posterior distribution. So an inferential approach to Bayesian statistics can be looked upon as part of the statistical decision theoretic approach. Inferential Bayesians argue that the posterior distribution describes the scientist's (within model) knowledge of a quantity. There is no need to do more. If and when that knowledge is needed, the full posterior distribution may be used in any calculations. Forming estimates, confidence intervals or whatever loses information, making later calculations less accurate. Nevertheless, convention, many journals and peer reviewers as well as the client scientist may demand precisely that. Moreover, accepting that Science is a value-driven activity, statistical decision theory does provide a mechanism for rendering explicit those values that are quantifiable, which is clearly important in some statistical analyses, e.g. experimental design, drug safety studies and more sensible approaches to hypothesis testing. Of course, the obvious compromise is to report both the posterior distribution and whatever estimates, confidence intervals and so forth that might be required.

Bayesian methods are usually characterised as being subjective. Many scientists and statisticians have found this of deep concern and, indeed, a large number still do. But all modelling and analysis involves subjective choices and judgements: what types of model to build and analyses to perform on them; what independence-dependence conditions are appropriate; is the evidence strong enough to draw a conclusion; and so on. Bayesian methods are just more explicit in some of their subjectivity, in particular using a prior distribution that injects – or, rather, can inject – prior knowledge into the analysis. Even then, *Objective Bayesian* methodology offers ways of minimising that prior knowledge so that the evidence accrued in the analysis comes almost entirely from the data (Berger, 2006; Consonni et al., 2018). Of course, some knowledge is also injected through the likelihood in Bayesian and non-Bayesian methods alike. For the present, I simply accept that Bayesian methods are subjective and proceed with the discussion on that basis. I shall return to the issue later in the paper.

My understanding of non-Bayesian statistical methodologies was shaped very strongly by Silvey (1975). Whereas Bayesians can work with probability distributions over the parameter space, Frequentist and several other methodologies deny the validity of such distributions[3]. So they are effectively only able to model variation and stochastic behaviours in the external world and, therefore, cannot use probability directly to answer questions about a parameter θ such as:

- What is the most likely value of θ?
- In what interval is θ likely to lie?
- Is it improbable that θ satisfies this hypothesis?

[3] Some, e.g. fiducial theory, did seek to offer non-subjective distributions over the parameter space, but I found them unconvincing.

Yet these questions are precisely the ones that interest a scientist, relating to estimation, confidence intervals and hypothesis testing, respectively. Silvey suggested that non-Bayesian approaches devised functions $g(X)$ of the data that display the data X in ways that are informative in answering these questions. For estimation, the function gives a value in the parameter space that has a high probability of being close to θ. For a 95% confidence intervals, there are two functions, $g_L(X)$ giving the lower bound of the interval and $g_U(X)$ the upper, such that 95% of the time the interval is predicted to include θ. For hypothesis testing, $g(X)$ is defined such that if the null hypothesis is true, its value will exceed some critical value only a small percentage of the time, conventionally 5% is to be 'significant' or 1% to be 'highly significant'. In these senses, $g(X)$ is *sensitive* to the scientific questions being asked. Before the data are collected, $g(X)$ is random with probability distribution derived from the sampling distribution of the data. After the observations are made, the data are known precisely, no uncertainty remains, and the function gives the estimate, confidence interval or hypothesis test outcome precisely. A reviewer of this paper has suggested that this perspective means that I see Frequentist methods as sensitive to the scientific question being addressed. I had not consciously been thinking in those terms, though the perspective is interesting. My concern in introducing $g(X)$ was not to make this point, but rather to suggest that non-Bayesian methods could be seen as presenting the data in an informative way in preparation for later discussions of exploratory data analysis (EDA).

4 The Process of Bayesian Statistical Analysis

Bayesians have always argued that their use of Bayes' Theorem and SEU was not to describe intuitive inference. Following the concerns of Edwards (1954), Kahneman, Tversky and others have produced abundant evidence that people's everyday behaviour was subject to many heuristics and biases relative to the Bayesian paradigm (Kahneman & Tversky, 1974; Montibeller & Winterfeldt, 2015; Morton & Fasolo, 2009; Tversky & Kahneman, 1981). Nowadays, one talks more of System 1 Thinking versus System 2 Thinking , the former referring to unconscious, intuitive thinking which may lead to dubious conclusions and the latter to explicit, conscious thinking within which principles of rationality may be adopted (Kahneman, 2011). This separation in two forms of thinking is, at best, simplistic, and there may a gradation of levels of consciousness between the two (de Neys, 2021; Evans & Stanovich, 2013; Shleifer, 2012). For our purposes, however, using the terms, System 1 and 2 Thinking is a useful shorthand for discussing how Bayesian analysis can support and nudge the processes of inference and decision towards the ideals of rationality encoded in its axiomatic basis, countering potentially misleading heuristics and biases (French, 2022; French & Rios Insua, 2000).

Before and during the 1970s, Bayesians drew sharp distinctions between normative and descriptive theories, emphasising that their analyses showed how people *should* infer and decide and not be overly concerned with how, when unsupported by analysis, they *did*. But during the 1980s, we began to recognise that our use of Bayesian rationality was prescriptive, i.e. it gained its value from both normative and descriptive insights (Bell et al., 1988). Focusing on statistical inference, prescriptive analysis advises scientists by analogy, showing them what an idealised model scientist (MS) should do in an idealised model world which parallels their own (French, 1986, 2022; Shafer, 1986). For a Bayesian, the MS is represented by the probability distributions and utility function. Mindful of possible System 1 Thinking, the analysis is embedded within a process of gently challenging discussions and deliberations to ensure that the scientists are not misled by any poor heuristics or biases. Further discussion of prescriptive decision analysis may be found in Edwards (2007), French et al. (2009), Keeney (1992a) and Gregory et al. (2013).

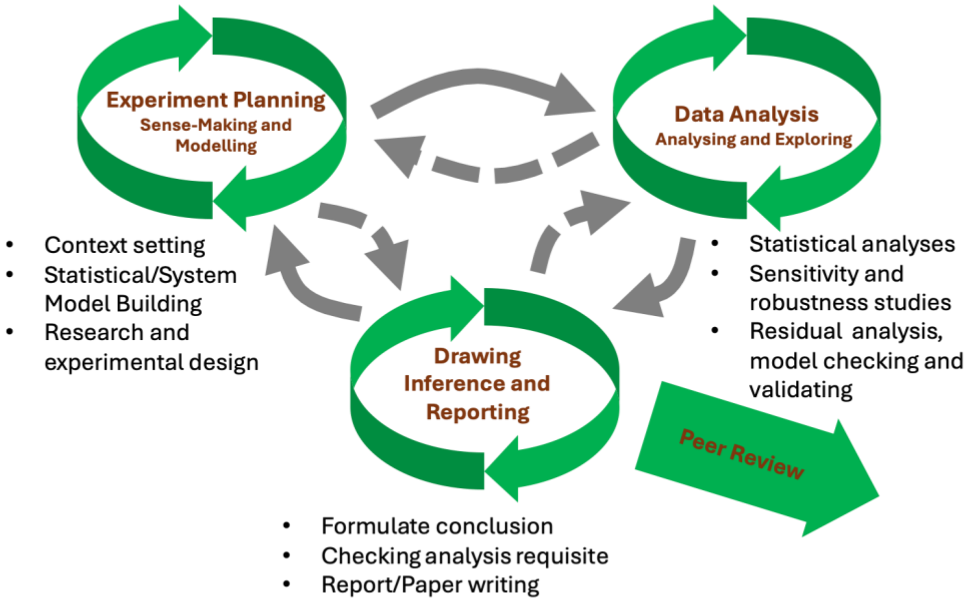

Figure 2: The Process of Bayesian Statistical Inference

Figure 2 further outlines this process of Bayesian statistical inference[4]. It begins with qualitative and, possibly, quantitative experimental design during which the statistician explores and makes sense of the situation, understanding its context and the broad set of research objectives, before focusing on the particulars of the problem and structuring a model of the relevant parts of the world and the MS, who embodies the knowledge that the real scientists wish to include in the analysis. Note that this knowledge might be the real scientists' actual beliefs, wider consensual scientific knowledge or, perhaps, an approximation to ignorance if they wish to be objective, letting the data 'speak for themselves' (French, 2024). During the next stage the model is analysed and explored to find and understand its implications, using Bayes' Theorem, possibly SEU, and various sensitivity and robustness techniques. The analysis indicates what the MS should infer from the data. The results should be validated using whatever model checks are possible, assessing the fit with all available data and performing residual analyses, other model checking and quantitative validation procedures. In the third stage, the real scientists make a judgement on whether the analysis has given them enough insight and understanding to make and report an inference. Is the analysis *requisite*, i.e. good enough, to move forward (French et al., 2009; Phillips, 1984)? This judgement will be driven both by the validation in the previous stage and also by comparing the inference against current scientific knowledge to see if it makes sense. However strong a statistical result is, it may still not represent the 'truth', i.e. explain the actual behaviour. So the scientists need to discuss its coherence with their wider understanding and knowledge. It may also be that aspects of the study have catalysed reflections that the model is lacking in some way or is too complex. Thus the team may decide that the model is not requisite, and the process needs to recycle. Indeed, at many points in the process, concerns may arise, and the process may iterate between or within stages. When the model is judged requisite, its implications have to be interpreted in real-world terms for reporting and sending to peer review.

The motivation of prescriptive analysis is that by observing and exploring the behaviour of the MS in a model world that parallels their own more complex world, the real scientists gain sufficient

[4] An idealised version in which the statistician is not simply presented with data!

understanding to decide what to infer: the model provides an informative analogy that allows them to do this, much as a religious parable informs moral thinking. The reason why I find non-Bayesian approaches unsatisfactory is that they do not construct such an analogy. Taking Silvey's interpretation, they simply transform the data in a way that should be informative to the scientists, but they leave the interpretation and understanding of that presentation to the scientists, thus possibly subject to misinterpretation through System 1 Thinking. Many investigations of heuristics and biases in intuitive inference suggest that 'irrationalities' are quite possible, though not inevitable: e.g. ignoring base rates or, in complex presentations, focusing on some aspects and missing others (Kahneman, 2011; Katsikopoulos, 2011; Miller, 1956).

5 Different Types of Uncertainty

Just over a century ago, (Knight, 1921) distinguished two types of uncertainty: unquantifiable or *strict uncertainty* versus quantifiable uncertainty or *risk*. Subsequent discussions developed various subcategories. Berkeley and Humphreys (1982) discussed seven types of uncertainty. In French (1995), I discussed ten types of uncertainty and now recognise that I missed some. There are many ways of categorising uncertainty and no real agreement on how to do so (for further discussion, see, e.g., Bradley & Dreschler, 2014; Granger Morgan & Henrion, 1990; Riesch, 2013). Over the course of many complex risk and decisions, I have come to appreciate many more forms of uncertainty that need to be attended to in any analysis (Constable et al., 2022; French, 2023; French et al., 2020). I find it useful to think to group these under three headings: roughly, those related to

- behaviours in the external world,
- judgements made in the process of modelling and analysing,
- lack of clarity, particularly about our values and objectives.

I will be pragmatic, seeking only to introduce the uncertainties that may arise in a statistical analysis and, more widely, in the process of scientific induction: a necessary preliminary step towards the more important question of how they should be dealt with.

5.1 Scientific Uncertainties about the External World

The world that provides a context for statistical, risk or decision analysis is inevitably an uncertain place. *Stochastic* (or *aleatory*) *uncertainty* relates to randomness in behaviour, variability in populations, and, of particular relevance to statistical inference, experimental errors. There is general agreement that stochastic uncertainties are well modelled by probabilities, be they subjective, frequentist or whatever. These uncertainties surround or maybe lie within a physical model of the world, in which there may be further stochastic and non-stochastic uncertainties. Current knowledge may provide the form of models to predict behaviours in the external world, but *parametric uncertainty* may need be dealt with before these can be used. Human and some animal behaviours are generally intentional and far from random, yet their unpredictability can lead to *actor uncertainties* for observers. There may be *one-off uncertainties* about whether a particular event will or will not take place. Bayesian statisticians accept all these as examples of *epistemic uncertainty* – i.e. a lack of knowledge or, more positively, a degree of belief – which can be modelled by subjective probability distributions. Non-Bayesians will find themselves less able to use probability to model some of these uncertainties; but, nonetheless, recognise the need to address such uncertainties.

There is a further uncertainty that relates to our uncertainty about behaviours in the external world: namely, *model uncertainty*. This describes how well the model through its structure and parameters can predict actual behaviour. There are ways of modelling a *portion* of this uncertainty using probabilistic methods. For early discussions see Blight and Ott (1975), Draper (1995) and French (1978), who approached the issue by adding an extra stochastic term to the model to inflate the overall uncertainty, terms that become smaller as the sophistication of the modelling becomes greater with increasing knowledge. More recent approaches of *reification* (Goldstein & Rougier, 2009), *conformal inference* (Gui et al., 2024; Lei & Candès, 2021) and *uncertainty quantification* (Berger & Smith, 2019) provide much more sophisticated approaches. However some aspects of modelling uncertainty are essentially unquantifiable, as we now discuss.

5.2 Modelling, Analysis and Process Uncertainties

In performing an analysis, there are some uncertainties that are often not well recognised; namely those that arise in selecting a methodology, choosing an algorithm including selecting any necessary approximations, how well modelling uncertainties have been modelled, the size of computational errors and deciding when the analysis is sufficiently deep or requisite. All these relate to the separation of the external world from the conceptual world (Figure 1). Attempts can and have been made to introduce to introduce quantified modelling of some of these uncertainties: e.g. for modelling errors as noted above, and for computational errors and approximations (Hennig et al., 2015; O'Hagan, 1992). But all relate to judgements on the part of the analyst and problem-owners reflecting the moves between the external and conceptual worlds (dashed lines in Figure 1). These judgements may be flawed, introducing error and uncertainty, and, as noted in Section 2, can never be fully modelled quantitatively. Comprehending the scale of error that might arise from choosing one method of analysis over another is extremely difficult. Judgements of whether a model and its analysis are requisite are, by definition, completely unmodellable, since modelling those judgemental errors begs the question of how good *that* modelling is and leads to an infinite regress. Analysts are invariably uncomfortable when challenged with the thought that their judgements, while leading to an analysis which will explore and reduce scientific uncertainties, will introduce such – hopefully smaller – *modelling*, *analysis,* and *process uncertainties*. it is a sanguine thought, and one that we should reflect on more than we do.

5.3 Internal Uncertainties about Ourselves

Risk owners, decision-makers and, we shall argue, scientists face further uncertainties arising not from lack of knowledge, but lack of clarity. Issues do not arrive on their doorstep ready packaged and immediately able to be modelled. Rather issues arrive as a mess of partially perceived entities and potential behaviours. The first task facing any problem owner is to clarify the issues and settle in their own mind what they need focus on. We shall discuss how they might do that in Section 7; but for now I would emphasise that this is just as true for scientists as for decision-makers. For instance, they need to be as explicit and precise as possible about their research objectives, both generally and for the specific experiment before them. Clarifying such issues is not something that can be sorted out by modelling and analysis, *per se*. It needs discussion and deliberation (French, 1995).

5.4 Deep Uncertainties

Knight's (1921) dichotomy between strict uncertainty and risk, which fell out of mainstream favour for many years, is again under active discussion, though strict uncertainties are now termed *deep uncertainties* (Marchau et al., 2019). Walker et al. (2013) define deep uncertainty as "the condition in which analysts do not know or the parties to a decision cannot agree upon (1) the appropriate models to describe interactions among a system's variables (2) the probability distributions to represent uncertainty about key parameters in the model, and/or (3) how to value the desirability of alternative outcomes." Moreover, there is neither the time nor resources to gather further evidence or resolve the disagreements before action is needed. Deep uncertainty may not be a major issue in statistical inference itself, because the pace of scientific induction is generally set by the speed with which data may be collected and analysed; but scientists and statisticians may, nonetheless, face such uncertainties. In 2020 when the Covid Pandemic struck, scientists and statisticians were called in by governments to advise on the massively deep uncertainties that were faced and to build models quickly to predict the disease's spread and impact. For a Bayesian perspective on how deep uncertainties may be dealt with in risk and decision analyses, see French (2022).

6 Cynefin: Exploratory and Confirmatory Inference

Taking a perspective from the growth of knowledge, and, thus as a corollary, the reduction of uncertainty, Snowden (2002) categorised four broad uncertainty categories using his Cynefin model. The Cynefin spaces describe the context for an analysis, categorising knowledge of cause and effect: see Figure 3. Note that I use "cause and effect" in the senses used by Snowden and others in the field of knowledge management, not with the technical precision found in many statistical discussions of causality (Dawid, 2002; Pearl et al., 2016). While modern statistical approaches to causal inference are very relevant techniques which might be used in scientific investigations, for our purposes so are any statistical methods which investigate relationships between entities.

In the *Known Space* relationships between cause and effect are well understood and all systems and behaviours can be fully modelled. The consequences of any action can be predicted with near certainty. Essentially, the only scientific uncertainties are stochastic; all others are small and can usually be ignored. In the *Knowable Space* cause and effect relationships are generally understood, but to model any specific behaviour further data are needed to address parametric uncertainty. In the *Complex Space* interacting causes and effects are poorly understood. Understanding of them is at best qualitative: there can be too many potential interactions to disentangle particular causes and effects. Thus quantitative modelling is difficult and often impossible. Contexts in the *Chaotic Space* involve events and behaviours beyond our current experience and there are no obvious candidates

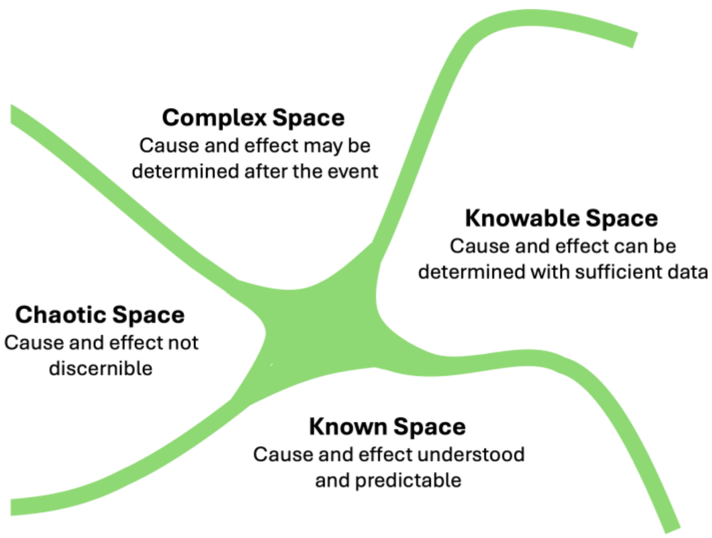

Figure 3: Cynefin

for cause and effect. Even separating and recognising entities can be difficult. Situations are entirely novel, and uncertainties are deep.

The central blob in Figure 3 is sometimes called the Disordered Space. It simply refers to those contexts that we have not had time to categorise. The Disordered Space and the Chaotic Space are far from the same. Contexts in the former may well lie in the Known, Knowable or Complex Spaces; we just need to recognise that they do. Those in the latter will be completely novel. One part of sense-making, which we will discuss in Section 7, is to assign contexts initially placed in the disordered to one of the other four. For further discussion of Cynefin: see French (2013, 2015).

A point to note is that only in the Knowable and Known Spaces is our knowledge of a situation sufficient to build sound quantitative models of what is happening. In the Chaotic and Complex Spaces, good physical models cannot exist since there is insufficient understanding of cause and effect.

Cynefin allows a straightforward, though ultimately somewhat naïve description of scientific induction. Suppose that scientists face a set of circumstances in which they has no understanding of cause and effect, a situation that was much more common centuries and millennia ago; though events such as the advent of BSE (‘mad cow’ disease) or the early signals of climate change provide modern examples. They perceive seemingly unconnected events and behaviours. In extreme cases, they might not be able to discern events, behaviours and even entities clearly. Gradually through interactions with their environment, they glean an understanding of some parts of it, seeing patterns or some such. Their perceptions of those parts of the environment move into the Complex Space as they begin to identify some potential causes and potential effects. At this stage, the scientists only perceive possible causes and effects. They have no clear idea of the scale of interactions between causes and effects and so cannot conduct controlled experiments. Investigations need to be exploratory. Through continued probing interactions, they make more sense of some of the behaviours and build a tighter understanding of cause and effect, beginning to build models and scientific laws. Confirmatory investigations now become possible. Although they still cannot predict perfectly, needing parameters or similar to complete the models; their understanding moves into the Knowable Space. Controlled experiments can now be designed to provide data to shape the models, further, determining parameters and so on. Some parameters may be found to be widely applicable, e.g. physical constants, and these aspects of scientific understanding moves into the Known Space, where effects can be predicted from a set of causes without further data.

Any particular process of learning is unlikely to move across all four spaces, instead making a small clockwise step in Figure 3, nor may that step necessarily take us from one space to another. Indeed, although conceptually possible, it is unlikely that any research project will begin in the Chaotic Space with behaviours barely perceived and understood and, from collecting and analysing a single data set, reach sufficiently precise inferences that the resulting understanding lies in the Known Space. Rather there will be a sequence of research projects, perhaps coordinated or perhaps independent, each of which may move understanding though Cynefin, ideally clockwise, though unanticipated or negative results may move it back anticlockwise. Few research projects start from a complete lack of understanding in the Chaotic Space; most start in the Complex or Knowable Spaces.

7 Soft Elicitation and Problem Formulation

One of the issues that I have observed is that statistics is often somewhat of an insular discipline. We develop our techniques and use them as ‘backroom analysts’, processing data through models. Although we often argue that scientists present us with poor or even useless data because they did not involve us in designing their experiment, both in the technical sense and more generally in whether it will truly inform their research objectives, we tend to focus on the statistical analysis. I believe strongly that this does us no favours. Through being involved in analyses within many research projects, we are uniquely positioned to observe and draw general conclusions about valid processes of scientific induction, and to use this knowledge to facilitate the entire process from the recognition of issues and the formulation of research objectives to the reporting of results and their assimilation into consensual scientific knowledge. Operational researchers, risk and decision analysts have long learnt to be much more involved in any project, seeing their role as much one of facilitation than just purely technical modelling and analysis (Eden & Radford, 1990; Edwards et al., 2007; Franco & Montibeller, 2010; Phillips & Phillips, 1993). In short, 'statisticians need to get out more!' Among the skills that we need to do this are those of soft elicitation.

Non-Bayesian statisticians seldom talk of elicitation; and when Bayesians do, they talk primarily of eliciting quantities: subjective probabilities, utilities or parameters. But elicitation should be much more than that, and in other professions it is (Checkland, 2013; French et al., 2009; Marttunen et al., 2017). It should relate to problem formulation. Indeed, models and their parameters are inseparable, so eliciting quantitative judgements without eliciting the wider qualitative knowledge necessary to build an appropriate model is only tackling half the job. I term the process of working with problem-owners to understand the context and issues and acquiring their knowledge, encoding what is learnt in a model or family of models as *soft elicitation* (French, 2021, 2024). *Hard elicitation* refers to eliciting quantities to populate these models (for a summary of such methods, see, e.g., Dias et al., 2018; Hanea et al., 2021). Note, however, I do not see a dichotomy between soft and hard elicitation, but rather a gradation of increasing precision between the qualitative and the quantitative (French, 2021).

That models do not arrive ready formed to conduct an analysis but need to be created is well accepted by analysts, though they may still reach for an off-the-shelf model rather too frequently. However, few textbooks acknowledge this, much less present methods to help in problem formulation. To be fair, any creative process – and this is a creative process – requires tacit skills, hard to render explicit. But the statistics literature could certainly do more; others have, especially that of operational research (French, 2021). I have often used the Cynefin Model to locate the context of an investigation, asking the clients about their relevant knowledge of cause and effect. In an investigation if such knowledge is meagre, I would expect the data analysis to be exploratory in nature; whereas more knowledge would suggest that confirmatory investigations and analysis might be needed. For further tools to shape statistical model-building, statisticians may point to Hand's (2020) *Dark Data*, which offers a series of tools/questions to identify data sets that are relevant to an investigation and ones that are not. More quantitatively, there is *exploratory data analysis* (EDA), famously promoted by Tukey (1977) and once conducted by pen and paper, but now largely superseded by scientific visualisation (Bendoly & Clark, 2016; Chambers, 1994; Zhang et al., 2010), machine learning and other AI-based methods (Hand et al., 2001; Klosgen & Zytkow, 2002; Rogers & Girolami, 2015), all of which can be used to explore preliminary data sets. Non-parametric methods offer a further step towards full parametric modelling and analysis.

The idea behind EDA is that the display of the data or some of its aspects can catalyse the recognition of a potential behaviour or effect. That is very similar to the interpretation that I hold of non-Bayesian methods. These redisplay the data in ways that are sensitive to the questions being asked by the scientists. Consider, for instance, regression and other multivariate methods. These redisplay the data in ways that emphasise potential interactions between putative causes and effects. The methods may be more complex than most EDA plots, but the intention to redisplay the data in informative ways is similar. I would not go so far as to believe in *p*-values as sound theoretical constructs (Wasserstein & Lazar, 2016), but these do point one to interactions that are seemingly worthy of further investigation. Thus, whereas I would not use non-Bayesian methods in prescriptive confirmatory statistical analyses, I have no great difficulty in adopting them as tools in exploratory analyses. I also note that non-Bayesian methods are similarly useful in residual analysis (Box, 1980).

One point that should be emphasised is recognising a potential pattern though EDA can be catalytic in driving an investigation forward, but it can also be misleading. Some patterns may be spurious and misleading. It is important that the statistician continually challenges the relevance of such patterns, asking the scientists to check them for coherence with current consensual scientific knowledge until evidence accumulates through subsequent analyses and investigations that they are indeed relevant.

EDA and related methods can only come into play once data have been collected and research objectives set. How does one decide what data to collect?[5] Belief networks and other graphical models are useful structures to help elicit knowledge of potential causes and effects and in turn suggest potential experiments or sources of data (Collazo et al., 2018; Oliver & Smith, 1990; Wilkerson & Smith, 2020). However, such techniques are, in turn, only useful after the entities and behaviours of interest have been identified and research objectives set. And that means that to build models, statisticians really need to explore the context and research issues much more widely.

There are many soft elicitation methods that can help catalyse discussion on these issues, encouraging clarity and rendering the broad context and research objectives explicit: e.g. soft systems (Checkland, 2013), cognitive mapping (Eden, 1988) and value-focused thinking (Keeney, 1992b, 2012). Rosenhead and Mingers (2001), Shaw et al. (2006), Shaw et al. (2007) and Smith and Shaw (2019) provide much fuller surveys of the methods. Simplistically, these are structured brainstorming methods which prompt discussion, capturing and relating ideas, displaying them from a variety of perspectives. Used carefully, they can lead to clear summaries of context and research objectives. I find the ideas of value-focused thinking particularly useful. Not only have I used them very successfully to elicit and structure research objectives, but I have also found them very useful in helping doctoral students plan and manage their research. Being clear on objectives helps maintain focus without being easily distracted by interesting, but probably irrelevant partial findings. Generally, I believe that statisticians focus too much on the model and technical analysis and not on the broader context of research and particularly on the research objectives. Without a good understanding of context, the softer, more qualitative aspects of validation may be missed, e.g. the coherence of the results with wider consensual scientific knowledge. Lack of clarity on research objectives can lead to poor choice of data sets and/or experimental design.

8 The Process of Scientific Induction

8.1 Introduction

In this section, I want to suggest a perspective on the entire process of scientific induction within a research project, arguing for a much more active role of a Bayesian statistician, facilitating deliberations, nudging the scientists continually into System 2 Thinking, away from unconscious biases. I believe that the statistician should be involved from the very beginning when the research

[5] Of course, today one does not decide what to collect. Technology collects so much data that the real question is often how does one decide what to select.

project is outlined, starting from the broader context and the initial formulation of research objectives and continuing through reporting to peer review and the inclusion of the findings into consensual scientific knowledge. In many respects my thinking has been influenced by the writings of George Box, Andrew Gelman and their colleagues (see, especially, Box, 1976; Box, 1980; Gelman, 2004; Gelman, 2018; Gelman & Hennig, 2017).

Drawing on my experience in prescriptive decision analysis, I see the role of the statistician as not just performing the quantitative analyses, but one of facilitating the whole process, continually challenging the scientists' understanding to ensure that they fully comprehend the conclusions that might be drawn and, perhaps more importantly, what ones cannot. Statisticians, because they are not driven by the specifics of the research imperatives driving the scientists and because of their somewhat greater career distance from the importance of the results, are ideally placed to facilitate discussions and deliberations between the scientists, and in the process can learn about the wider context and research objectives themselves, improving the value and validity of the models that they build and the analyses that they conduct.

8.2 Overview

Figure 4**Error! Reference source not found.** provides an undoubtedly idealised overview of the process of scientific induction. Clearly, embedded within it is that of statistical inference (Figure 2), surrounded by a much more comprehensive first stage involving the early formulation of research issues before research objectives are fully clarified and an extended third stage covering peer review and publication of the scientific results, the archiving of data and analyses, incorporation of the conclusions into consensual scientific knowledge and general communications to wider stakeholder groups. It is an iterative process with much reflection on and checking of earlier steps, one that may cycle several times as understanding grows and scientific hypotheses are refined. I believe that the statistician's role should be extended to facilitate this longer process; though undoubtedly some initial discussion will take place between the scientists, before they invite the statistician to join the discussions and planning.

The process may begin in any of the Cynefin Spaces. If it begins in the Chaotic or Complex Spaces, then the research will be exploratory in nature, seeking to understand what entities are involved and how their behaviours may be related. Thus appropriate statistical methodology will lean heavily towards EDA and related techniques. If these provide the scientists with enough insight to perceive potential cause and effect behaviours that can be modelled, then confirmatory statistical methods may be used, but probably with the proverbial pinch of salt. At such an early stage of understanding, the modelling uncertainties will be high and are unlikely to be captured fully, if at all, within the analysis. If the research project begins in the Knowable or Known Space, then the scientists' understanding of cause and effect will be sufficient that physical models can be structured and the investigation will be more confirmatory in nature. Confirmatory statistical methods will be the main methods of analysis, though EDA will still be needed in relation to data checking and later residual analyses.

When the scientists have deliberated on the results of the analyses, they will need to judge whether the evidence is sufficient – requisite – to put their inferences, whether positive or negative, forward for peer review. If they believe that it is, they will need to write and submit reports. If not, then their research project will cycle back to plan further experiments, data collection and analyses; as they also will if peer review finds their inferences insufficiently supported. Publication will lead to the inferences being assimilated into consensual scientific knowledge, ideally after they are checked

To be read to the Royal Statistical Society at the RSS Annual Conference, at 9am on 9th September 2026, Bournemouth International Centre

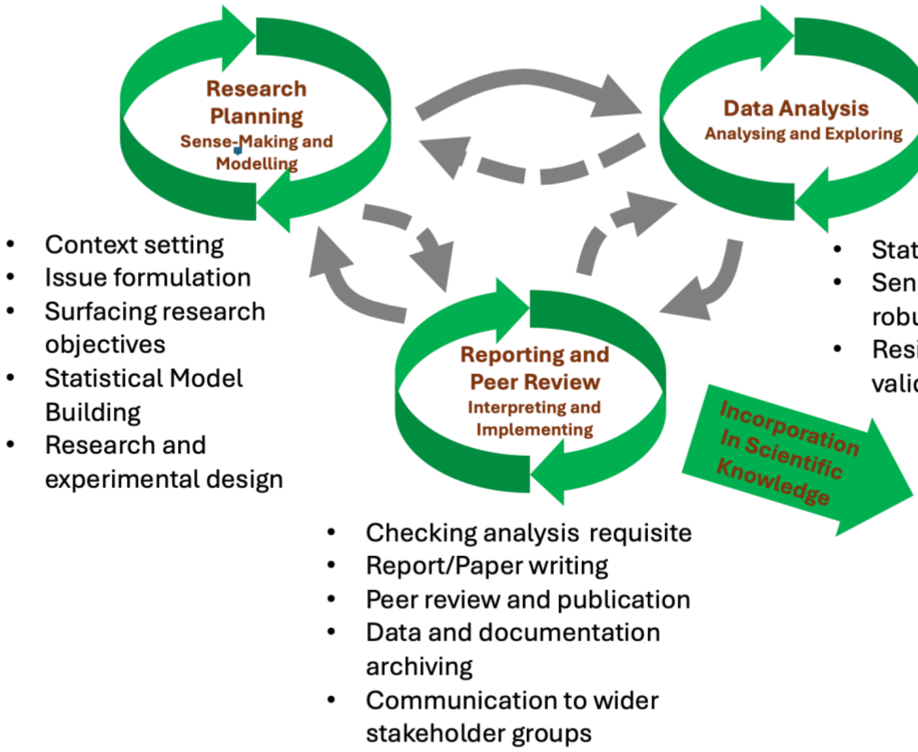

Figure 4: The Process of Scientific Induction

through independent studies, either through replication studies or alternative experiments and data collection that explore the same cause and effect behaviours by different means.

Below we discuss each of these stages in greater detail.

8.3 Research Planning

A research project needs to be carefully planned if it is to be effective, addressing feasible and useful research questions. It needs to begin with sense-making, with the aim of setting the context and understanding current scientific knowledge of the topic as well as surfacing the scientists' own views and hypotheses. I use 'hypothesis' here in the general sense of expectations about what the research may find, not a tightly phrased statistical hypothesis. Soft elicitation discussions that seek to locate the research issues on the Cynefin Model can be very informative. If current knowledge lies in the Chaotic or Complex Spaces with little understanding of what might be happening, these steps may involve much reflection and discussion; in the Known or Knowable Spaces, these may be simpler, but still necessary. It is particularly important to clarify research objectives, being explicit about whether the research is exploratory or confirmatory, so that an appropriate experiments, data collection and statistical analysis strategies can be set. In exploratory research, soft elicitation tools will help catalyse thinking and shape the data collection. These may range from seeking to develop a cognitive map, qualitative belief net or chain event graph to using Hand's (2020) categories of dark data to help identify experiments, surveys or other sources of existing data. In the case of confirmatory research, both soft and hard elicitation will be needed to build a putative physical model of the phenomenon and a statistical model of the observational error structures. The data collection strategy should cover not just the observation process itself, but data storage in preparation both for the analysis to follow and subsequent reporting and archiving.

There will be a need to deliberate on the form and details of the model scientist (MS), particularly the prior distribution. The MS's prior might be based on the lead scientist's own prior knowledge or perhaps a consensus view built from the team's knowledge, but this is not that common. It risks bias arising from the team's hope for a positive impactful result overweighting the actual evidence from the experiment. The team or perhaps the statistician might scour the literature to identify consensual scientific knowledge that is accepted by the wider community and then use this as prior knowledge: e.g. one of my first, and arguably most successful, forays into Bayesian statistical analysis used an informative prior, relating to well-established theoretical and empirical knowledge of distributions of crystallographic data (French & Wilson, 1978). Sometimes a prior is designed to favour simpler models, embodying not specific knowledge of the context, but Occam's Razor: namely that a simpler explanation is generally preferred to a more complex one, simpler models being easier to verify and more robust in use (Jeffreys, 1961). On the other hand, the scientists, striving to be objective, may not wish that any prior knowledge be drawn into the MS's prior and hence the analysis. In that case, there are prescriptions within the Objective Bayesian community of possible uninformative priors, embodying some approximation to ignorance so that the data 'may speak for themselves' (Berger, 2006; Consonni et al., 2018). Note, however, that in any analysis some prior knowledge is drawn in through the physical and statistical modelling of the surrounding world and in defining the parameter space and any dependence/independence conditions separating structures in the models. Further in capturing the scientists' wishes for what prior knowledge should be drawn into the analysis, the statistician may choose or construct a form for the prior distribution which has nice computational properties: e.g. a conjugate prior or one that helps MCMC convergence. Thus the prior distribution may not represent the (full) prior knowledge of any of the investigating scientists; and, indeed, may be an amalgam of several real and imaginary perspectives: see French (2024) for further discussion

Although together they set the context and objectives of research, both generally and for the particular investigation, these formulation and planning processes are seldom organised into a structured meeting, not at least until there is a research funding application form to be completed. If such a meeting is organised, it is seldom facilitated. Yet such deliberations are usually much more effective if facilitated by a relatively independent person, who can nudge individuals away from System 1 Thinking and potential biases and can also intervene if tendencies to aberrant group behaviours such as group think become apparent (Phillips, 2011; Phillips & Phillips, 1993; Reagan-Cirincione, 1994). Here I am suggesting that the statistician takes on this role. Many Bayesian statisticians are already skilled in hard elicitation of probabilities, utilities and parameters from groups (Dias et al., 2018; Hanea et al., 2021; O'Hagan, 2019). Extending their skill set to facilitation of the much softer early deliberations that would be needed here is not a large step. In my experience, such facilitated soft elicitation meetings have huge benefits, particularly in setting research objectives which set the frame for experiments, data collection and the subsequent analysis. Moreover, facilitating such meetings provides the analyst with considerable insight into the research, its context and importance, and above all what it is intended to achieve. That insight enables much more sensible analytical judgements on experimental design, methods of analysis and so forth, hopefully reducing the modelling, judgemental and computational errors. Finally, such sense-making and problem formulations processes define the 'language of the project', creating a clear terminology in which to frame the issues and describe the behaviour under investigation. That does much to ensure clear and well drafted reports at the project's conclusion.

8.4 Data Analysis

When the data have been collected, they need to be analysed according to the strategies developed during the planning discussions. The first step is to use EDA to check that the data collection has proceeded as planned and whether anything untoward stands out in them. This is, of course, an obvious preliminary for exploratory research before other tools such as multivariate analysis, regression and machine learning are used. But for confirmatory research, there is a question about what to do if something unexpected is discovered. Should the analysis proceed as planned? Should it be modified? Should 'outliers' be included or deleted. What should be done about missing or lost data, particularly if there are systematic losses? I favour returning to the research planning stage and restructuring the analysis, even perhaps considering collecting further data. However, my experience in statistical analysis has tended to be in 'pure' science and not in areas in which the results will feed into regulation. In some research, e.g. medical and pharmaceutical studies, it is a requirement that data collection and analysis protocols are agreed in advance and followed exactly. The philosophy behind Frequentist approaches requires this since the observational distribution needs to be defined and fixed before the data are collected. The protocols can, of course, contain contingency plans for dealing with different types of missing data, outliers, etc., but they should not be changed. Some Bayesians may make the same argument.

For an exploratory study, the statistician can use many EDA, machine learning or, following my arguments above, non-Bayesian methods to explore the data and see if any patterns stand out. Gelman (2003; 2004) discusses Bayesian approaches to EDA. The statistician will have been primed by the earlier discussions with the scientists on context, current understandings and potential causes and effects to have at least some idea of what an interesting pattern might be. But the statistician is not one of the scientists. It is imperative that the scientists also explore the data, ideally without prompting from the statistician to look at particular features. During confirmatory studies, Science may need to be 'objective' in some sense. But in the early exploratory analyses in the Chaotic and Complex Spaces, creative insight is needed to suggest how possible patterns may make sense as possible cause and effect behaviours. The purpose of exploratory techniques is to catalyse such insights. For that reason, I strongly favour a meeting between the statistician and the scientists which takes the format of a decision conference and which the statistician facilitates (French et al., 2009; Phillips, 2007). Modern statistical systems such as R mean that the exploratory analyses can be conducted live during such meetings allowing the scientists to see and drive the explorations. The statistician need not – and should not – prompt the meeting with personal observations, unless these are completely missed by the scientists. In such cases, these observations , along with the statistician's wider understanding of uncertainties and spurious patterns, provide a basis to gently challenge the scientists thinking. Apart from helping the scientists to develop their own informed insights, this process also ensures that they comprehend the data far better than if they are simply presented with a report or slideshow of analyses conducted elsewhere by the statistician.

In a confirmatory study, if the initial EDA suggests some unexpected issues or problems with the data collection, the statistician might need to return to the scientists to discuss how to proceed; otherwise, the analysis may proceed without involving the scientists, who are unlikely want to be involved in, e.g., tuning the convergence of an MCMC calculation or similar computations. Whatever the case, the analysis should not just be accepted, but subject to sensitivity and robustness studies to check for the risk of spurious results arising from some flaw in the model building or computational design. There is a need to explore residuals between the model and the data to investigate possible systematic deviations, which might indicate some behaviour that is missing in the model. Residual analysis and EDA were developed contemporarily and share many techniques (Freund et al., 1961; Martin et al., 2017). Note that Frequentist methods may also be used in residual checking for Bayesian analyses. Box (1980) essentially argued this, noting that the analyst, when stepping out of the model and back to the real world, needs to criticise and challenge the conclusions with the data and other knowledge not necessarily encoded in the model (See also Bayarri & Berger, 2004; Gelman, 2004). As with EDA, residual analysis should be curiosity-driven, seeking patterns that could represent systematic behaviours. The statistician should ensure that the scientists explore and see a full analysis conducted to professional statistical standards. Again there are advantages to presenting these results in a facilitated meeting with many of the analyses conducted interactively with the scientists to answer questions such as: "What happens if X and Y are not assumed to be independent?" or "Should we drop that outlying observation?" At the end of this process, the scientists should understand what inference the MS is making in the model world. Indeed, ideally, using sensitivity analysis on the prior, they should see what each member of a family of MSs would infer, which will provide them with a understanding of the robustness of the results (French, 2003). These related analogies will help the scientists draw their own inferences in the real as opposed to model world, helping them avoid any System 1 Thinking issues.

Confirmatory studies need to be validated in several ways. Within model checks, residual analysis, etc. can indicate a flawed analysis or poor fit of the model (see, e.g., Gabry et al., 2019; Gelman et al., 2020). But formal statistical analysis is only part of the process of scientific induction (Hubbard et al., 2019). There needs to be an assessment whether the results cohere with wider consensual scientific knowledge. Are they surprising in any way? Some scientific results are undoubtedly surprising, but the majority simply represent an incremental growth in knowledge that builds on current understanding. If an inference sits uncomfortably with current knowledge, it should be checked carefully and only taken forward for peer review if the evidence is sufficient. This wider check on validity is one of the reasons why the statistician should explore the context and wider consensual knowledge during research planning. Being aware of that background gives him or her the ability to make a preliminary judgement, one that can be tested in discussion with the scientists in the next stage.

8.5 Reporting and Peer Review

The next stage begins with further discussion of and a decision on whether the statistical analysis, model checking and quantitative validation procedures have provided the scientists with sufficient evidence to move forward to an inference. In their deliberation, *all* uncertainties should be considered, not just those amenable to quantification. Are the scientists content that the judgements made on independence, distributions, analytic and computational methods are reasonable? If EDA led to changes in a planned confirmatory analysis and so, in a sense, some data was double-counted, care will be needed here. From our perspective that Bayesian analysis provides an idealised model of inference, it simply means that a 'larger pinch of salt' will be needed in making the induction by analogy. How the scientists learn from the analysis will be moderated by how closely they feel the idealised model of inference reflects the actual situation. When the analysis has proceeded as planned, they may feel a little more confident in drawing their conclusions, as may subsequent peer review. When something has led to a modification of the plans, they should be less confident, but only rarely should they reject the results in their entirety (see also Gelman, 2003; Good, 1983). As noted above, the scientists also need to compare the inference against current scientific knowledge to see if it makes sense. Knowledge should be coherent. If it does not cohere in this soft sense, but nonetheless the quantitative analysis can find no clear fault with the 'fit' of the model, then some reflection will be necessary. Negative and discordant results should undoubtedly be published, but the evidence must be strong and the team must be convinced that their judgements are sound and that all uncertainties have been adequately discussed. The team may decide that the model is not requisite, and the process recycles. Throughout these deliberations, the statistician should drive the discussion with the scientists, gently and continually challenging their statements in ways that nudge them out of any System 1 Thinking into explicit System 2 Thinking.

If the analysis is judged requisite, the process moves into interpreting the results and their relevance to current scientific knowledge, writing reports and sending them for peer review and publication. As they do this, they should also ensure that their archives are structured and completed as planned. During the reporting phase, the statistician should continue to check the scientists' understanding to ensure that they truly comprehend what the results mean and are not misled by any foibles of System 1 Thinking. Modern publication practices are influenced by the needs of career progression and research grant applications as much as the communication of results. Negative outcomes of research are not welcome, a further potential to bias scientists, at least subconsciously, towards more positive, but less justified interpretations of their results.

Good reporting and sound peer review are as much guardians of 'scientific objectivity' as replicability: see Gelman and Hennig (2017). Strictly, good peer review should be much more than reading a paper submitted for publication and checking whether the results make sense, it should check the analysis with some care. For that, access to the data and collection protocols, other documentation and logs of the process are needed. Fifty years ago, employers of the wider peer community in universities and research institutes recognised this and allowed time for such reviews. Sadly that is far from true these days, despite current concerns about replication (Boulesteix et al., 2020; Carriquiry et al., 2023; National Academies of Sciences, 2019; Romero, 2019). If the result is important enough for full replication studies, then such archives are even more essential. As already noted, documentation which discusses *all* the uncertainties inherent in the analysis and inference are particularly important; though, of course, such discussion should be in the published papers, not just the archives (Schwab et al., 2022).

I emphasise again that the process of scientific induction may iterate both within stages and between them as aspects of the discussion or analysis create an awareness that some early assumption or conclusion may be flawed. It may be that no useful results are found, other than that the expectations and hypotheses that drove the research are unsubstantiated. But that *is* the process of scientific research. Often it is necessary to go around the full loop through three stages several times. This will almost invariably be the case, if a research team's investigation begins in the Chaotic Space, when a phenomenon is barely perceived. Moreover, I am writing as if *one* team of scientists conduct the research; generally there are several teams investigating the phenomenon. In pure science, they will be share ideas and results; in the commercial world, they may be working independently to protect intellectual property rights until they can be secured by patents. Science, in general, progresses by multiple transitions of the loops in Figure 4**Error! Reference source not found.**.

9 Some Remarks on Reporting

9.1 Impartial Honest Reporting

I have acknowledged my subjective philosophy already, but I acknowledge that many feel a need to be objective in some sense. Not perhaps, in terms of some absolute, objective, unarguable perspective on the world, but in terms of a wish to be impartial and unbiased. Indeed, that is how the

Oxford English Dictionary (OED) defines objectivity[6]: "The quality or character of being objective; (in later use) esp. the ability to consider or represent facts, information, etc., without being influenced by personal feelings or opinions; impartiality; detachment." It is in that sense that I wish to discuss objective scientific reporting. I believe that much of current scientific reporting and particularly that of statistical analyses fails to report the full uncertainty in the results, albeit unintentionally. Too often the uncertainty associated with an inference is stated as a subjective probability distribution, confidence interval or simply a standard deviation. There are inevitably non-quantifiably uncertainties that remain and need to be discussed in order to set the quantifiable ones in a wider context, even when a Bayesian analysis has led to a full (within model) posterior distribution.

9.2 Research Objectives within a Value-Informed Context

In presenting soft elicitation and sense-making, I emphasised the use of value-focused thinking in setting clear research objectives. Clearly any report should state these objectives, but it should also discuss their context and how they led to the choices made in designing the data collection and analysis, particularly in setting any loss function used. The discussion should recognise when any value judgements led to specific assumptions that limited or might have limited the analysis in some way. The gentle challenging within facilitated soft elicitation should have made explicit many assumptions and led to a more impartial report than would otherwise have been the case.

9.3 Within Model Uncertainties

Because Bayesians have the full armoury of subjective probability theory to model uncertainties within their models, they can analyse and explore the scientific and, perhaps, computational uncertainties associated with the inference. The posterior distribution itself forms a comprehensive summary of how these uncertainties are balanced in the inference that the MS would make in the model world. The advent of graphical computing has meant that there are now many ways in which these modelled uncertainties can be visualised and presented in reports (Chen et al., 2008). Robustness and sensitivity analysis techniques are also important for exploring the uncertainties within models (French, 2003; Rios Insua & Ruggeri, 2000): e.g. independence conditions can be changed and the effect on the model inference observed. Particularly important are explorations of varying the prior distribution used. This both shows the effect of the prior in moderating what the data alone would suggest as an inference and also allows the readers to see the inference to which using their own knowledge in the prior would lead. Obviously, the case for adding the conclusion to consensual scientific knowledge is stronger, if relatively large variations in the prior from that used in the analysis lead to much the same inference.

9.4 Unmodelled Non-Quantifiable Uncertainties

However, I contend that no Bayesian analysis, indeed, no analysis can model all the uncertainties or analyse them fully. For instance, consider modelling uncertainties (Sections 5.1 and 5.2). Although it is possible to model a proportion of these, doing so cannot deal with all modelling error, because the techniques introduce their own modelling errors, hopefully smaller, but present nevertheless. Judgemental errors introduced by the choice of data collection protocols, modelling strategies, analytic techniques, etc. might be treated in the same fashion, but some judgemental uncertainty will similarly remain. Then there is the specific decision related whether the model and analysis is requisite. Judgemental uncertainty related to that cannot conceptually be modelled, because it is a

[6] oed.com, visited 2nd September 2024.

decision to cease modelling entirely. So there will be some unmodelled uncertainties, unmodelled because they are essentially unquantifiable. I am well aware that many statisticians, indeed many analysts across many disciplines are uncomfortable with this statement, but models exist in our conceptual world not the external world. It is the external world that scientists investigate to understand and anticipate behaviours 'out there'. If we do not try to provide information on the full uncertainty on those behaviours, then our understanding of those behaviours and any predictions that we may make may be overconfident. Moreover, it will difficult to check whether any future replication coheres with the results of the study.

Kuhn (1961) argued, in effect, that good scientific reporting should instil across the scientific community a tacit understanding of the likely correspondence of current knowledge with reality. But he did not suggest that it could be quantified perfectly. Tacit understandings are exactly that: an intuitive feeling not easily rendered explicit, not easily shared (Boisot, 1998; Nonaka, 1991; Polyani, 1962). So it will not be easy to report and convey the remaining, unmodelled uncertainties. But that does not mean that we should not try; not to do so would be an abrogation of our scientific responsibilities, our objectivity in the sense of the OED definition.

So, while such uncertainties cannot be quantified, they can be discussed. They arise from the many judgements that structure models and an analysis. As a first step such judgements should be acknowledged, and often they are. The reasoning behind all the assumptions, analytic techniques, and wider contextual validation of the inference against current knowledge should be given. Readers and peer reviewers can then better assess if they would have made the same judgements as well as that of requisiteness. The scientists might also use their final (posterior) model to predict future behaviours and publish the results tabulated against what actually occurs, not as a validation test *per se*, but to instil a feeling of how the predictions pan out against the actual behaviour. However, I do not expect any single paper or study to instil a feeling for the uncertainty in making such predictions. Rather I would argue that if it became the norm to report scientific investigations with such discussions and exemplar predictions, then the community would build up expertise in interpreting them with 'appropriate pinches of salt'. The literature on knowledge and knowledge management suggest that tacit understandings are built and shared though experience rather than any single discussion in any single report (Boisot, 1998; Cianciolo et al., 2006; Nonaka, 1991; Polyani, 1962).

9.5 The Replication Crisis

The current Replication Crisis in Science speaks, in part, to the fact that scientific papers and communications do not report investigations satisfactorily. There are undoubtedly some deliberately false reports of experiments; but I believe that most researchers are honest, only lacking an adequate understanding of the uncertainties inherent in their conclusions. Indeed, I believe this to be particularly true when professional statisticians have not been involved in the entire process as is often the case across all the sciences, but perhaps particularly social sciences. The researchers themselves, guided by the demands of their disciplinary journals, may simply have applied analytic techniques they acquired from undergraduate courses in "Statistics for subject X" or similarly titled textbooks, reporting standard quantifications of the uncertainty. Unmodelled, unquantified uncertainties are unrecognised and ignored, and certainly not discussed and reported in the senses that I am suggesting here. So at least some of the replication failures currently surfacing may be related to such underreported[7] uncertainties. Indeed, investigating some of the replication studies, both those that

[7] Underreported in either published papers or associated data and experimental archives.

failed and those that successfully replicated the original results, might help instil that feeling of confidence in the correspondence between model predictions and actual behaviour discussed by Kuhn (1961).

9.6 Communication to a Wider Community

Good science needs to be communicated to the wider public, not just to fire their interest and to show the value of investment in research. Among the public are many other scientists, innovators and engineers in other disciplines. Even the most esoteric result in some area of fundamental or theoretical research can trigger ideas in other disciplines, developing new areas of research or a technological advance: e.g. obscure number theory has revolutionised cryptography. So the research project team should also develop a public communication strategy for all, but the most meagre of incremental advances. Such communications should open and honest about *all* the uncertainties inherent in the research. The public perceive that that scientists are much more certain in the results of their investigations than they actually are. So when scientists make predictions relating to, say, some risk, those predictions are taken as much more accurate than they are likely to be; and, if those predictions turn out to be wrong, the public become confused and lose confidence in the risk management strategy (Bennett et al., 2010; Lord Phillips, 2000). It is important to educate the public to understand the true uncertainty in scientific knowledge.

10 Discussion

10.1 Of Course, It's Not Like That…

This description of the process of scientific induction and the role statistical inference within it is clearly an ideal and, sadly, one seldom fully applied in practice. Many of us whose jobs have included the support of research data analysis have often despaired when a knock at the door heralds the arrival of someone who says: "I have some data. Could you analyse it?" Such knocks on the door have thankfully reduced in frequency over my career, but alas not to a rarity. It is still the case, particularly in academia, that researchers do not understand that statistical analysis should be considered from the earliest planning stages of a project through to its reporting. If it is seen as an 'add-on', in the worst cases simply as something demanded by the publication requirements of journals, then potential disaster awaits. Too often I have had to explain that the quantity and quality of the data collected cannot inform the research objectives, if these ever have been articulated clearly. Perhaps worse, in many research projects advice is never sought from a statistician. Instead one of the more numerate of the research team applies rote statistical analyses to their data using one of the many user-friendly packages available. Recently I received notice of a seminar describing how the R statistical system could be embedded in ChatGPT to provide an AI system that would obviate the need for scientists to consult anyone with statistical knowledge. I despair! Good scientific research does not need less interactions with professional statistical analysts: it needs more. Statisticians with their specific perspective on inference rather than contextual knowledge could and should contribute more widely to the broader research process than many currently do. See Gelman (2018) and Stark and Saltelli (2018) for (slightly) less polemical arguments along these lines.

10.2 The Statistician as a Facilitator

Within risk and decision analysis, it is common to make a distinction between contextual and procedural knowledge. The problem owners, their experts and stakeholders bring contextual knowledge. They understand the issues faced. The analysis team bring process knowledge. They understand how to support risk management and decision-making using modelling and analysis and so can facilitate the process effectively as well as conduct the technical calculations. They understand prescriptive support. I am suggesting that this distinction between contextual and process knowledge needs better recognition in statistics and that statisticians take on the role of facilitation in scientific research. Many statisticians may not welcome this proposal, even if they accept the arguments earlier in this paper. Perhaps I should separate *applied* statisticians from *mathematical* statisticians, who are more driven by an interest in the theory and techniques of inference and computation, and do not want to acquire and deploy the softer skills of facilitation. I should admit that much the same is the case for some risk and decision analysts. So in many cases, teams of facilitators and analysts support a project, while in others individuals perform both tasks. That can obviously happen within scientific induction too; though to those who have not tried facilitation, I would advise them to try it as provides much clarity and insight which is both useful in conducting the analysis within the project and more widely in developing inferential methods[8].

10.3 Bayesian Analysis as a Model of Inference

In Section 3, I quoted Lindley's summary of three key assumptions in using Bayes' Theorem and SEU theory as a normative model within a prescriptive analysis. Much of this paper has sought to set these in context and recognise that in practical inference there are issues with making each of the assumptions. In the case of his assumption (a), I have argued that it is not the case that all uncertainties can be modelled by probability. For deep uncertainties, it might be that conceptually they can be modelled probabilistically, but not usefully in the time available for the analysis. And then there are some uncertainties that are non-quantifiable, e.g. uncertainties relating to the choice of modelling assumptions or analytic techniques. Such uncertainties relate to those aspects of reality that omitted in the model and its analysis. They lead to unmodelled and unquantified uncertainties that remain when the analysis is concluded. I find Lindley's assumption (b) that the worth of an entity or outcome should be measured by its utility much less controversial. However, Bayesian statistical analyses often ignore it, simply because they focus on inference and report posterior distributions. When such analyses do consider worth or, more usually, the loss from an error, they tend to do so in rather simplistic, conventional ways. I have little problem with his assumption (c). It is the SEU maximisation that is the driving model in risk and decision analysis though it is far from easy to achieve in some cases (French, 2023; French & Argyris, 2018).

With those caveats, I have argued that the Bayesian model of inference and decision sits at the heart of prescriptive analyses. In statistical inference, the MS's inferences inform scientists by analogy showing them what a rational scientist would infer from which represent model observations which represent their data. In doing so the process nudges them into explicit System 2 Thinking avoiding possible pitfalls associated with unconscious System 1 Thinking. Non-Bayesian techniques do not provide a model of inference, leaving the scientists to make their inferences intuitively based on a

[8] It can also be great fun!

function of the data which is informative in relation to the research question being asked, thus offering a greater risk of System 1 Thinking errors.

10.4 Institutional and Cultural Issues

There are many institutional and cultural barriers to be overcome before the vision of statistical involvement in scientific induction suggested here can become regular practice. Professional bodies and educational institutions need to agree with these ideas and introduce them into their recommended training syllabuses and accreditation programmes. The wider science community need to consider and welcome them. Culturally, both statisticians and researchers need to adopt research practices with much more interaction between their communities; and that, dare I say, will cost more. Although there are many more Bayesians in the statistical community than when I began my career, they are not an overwhelming majority. In other disciplines, Bayesian methodologies have made fewer inroads. Many journals in the empirical disciplines impose page length restrictions on papers, effectively curtailing any discussion of the uncertainty in the results beyond standardised measure of statistical confidence or significance. I could write for pages on the barriers to the adoption of these ideas. But that does not mean that the debate is not worth having. Stirring up our community to think more about *all* the uncertainties faced by a scientific investigation, the need to report them better and the need to guard against System 1 Thinking errors can only improve our practice in the short term. Moreover, individual committed Bayesians can adopt the ideas here whenever possible, inveigling themselves into the full process of scientific induction. I have witnessed huge progress in the practice of statistical inference in the past 50 years: I do not believe that it will stop in the next 50. So perhaps some of the ideas here can shape the direction that progress will take.

11 Conclusions

I began with four objectives.

i. On the role of statisticians, I have put much emphasis on what their process expertise can bring to scientific induction and suggested that a much stronger role in facilitating the process with the added benefit of much greater understanding of context and hence the potential to build better statistical models and analyses.

ii. Scientists are human and subject to possible unconscious biases. I have argued that using the latest prescriptive methods would allow statisticians to counter such biases, nudging scientists towards more explicit and rational thinking in their inferences contributing to better scientific induction.

iii. I have discussed the many uncertainties involved in an investigation, suggesting that some are ignored in most statistical analyses and scientific papers which may explain some of the current concerns about replication.

iv. Good and objective reporting is important to ensure that consensual scientific knowledge grows soundly and that the users of the scientific results comprehend the remaining uncertainties and their underpinning assumptions.

To be read to the Royal Statistical Society at the RSS Annual Conference, at 9am on 9th September 2026, Bournemouth International Centre

Acknowledgements

I am grateful to very many people who have discussed Bayesian ideas over my career, including those who in no way supported them. This synthesis has been shaped by many such discussions. The reviewers of earlier versions of this paper made some very pertinent suggestions. I am also grateful for many grants and research contracts from UK, European and International research bodies and agencies. To all, thank you. No specific funding has supported this work however, and I take full responsibility for the ideas herein.

References

Barnett, V. (1999). *Comparative Statistical Inference* (3rd ed.). John Wiley and Sons.

Bayarri, M. J., & Berger, J. O. (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science*, *19*(1), 58-80.

Bell, D. E., Raiffa, H., & Tversky, A. (Eds.). (1988). *Decision Making: Descriptive, Normative and Prescriptive interactions*. Cambridge University Press.

Bendoly, E., & Clark, S. (2016). *Visual analytics for management: translational science and applications in practice*. Taylor & Francis.

Bennett, P. G., Calman, K. C., Curtis, S., & Fischbacher-Smith, D. (Eds.). (2010). *Risk Communication and Public Health. 2nd Edition.* Oxford University Press.

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*(3), 385-402.

Berger, J. O., & Smith, L. A. (2019). On the statistical formalism of uncertainty quantification. *Annual Review of Statistics and its Application*, *6*, 433-460.

Berkeley, D., & Humphreys, P. C. (1982). Structuring decision problems and the 'bias heuristic'. *Psychological Bulletin*, *50*, 201-252.

Blight, B., & Ott, L. (1975). A Bayesian approach to model inadequacy for polynomial regression. *Biometrika*, *62*(1), 79-88.

Boisot, M. (1998). *Knowledge Assets: Securing Competitive Advantage in the Information Economy*. Oxford University Press.

Boulesteix, A. L., Hoffmann, S., Charlton, A., & Seibold, H. (2020). A replication crisis in methodological research? *Significance*, *17*(5), 18-21.

Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791-799.

Box, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society*, *A143*(4), 383-430.

Box, G. E. P., & Taio, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley.

Bradley, R., & Dreschler, M. (2014). Types of Uncertainty. *Erkenn*, *79*(1225-1248).

Carriquiry, A. L., Daniels, M. J., & Reid, N. (2023). Special Issue on Reproducibility and Replication. *Statistical Science*, *38*(4).

Chambers, J. M. (1994). *The Elements of Graphing Data*. Murray Hill, New Jersey.

Checkland, P. (2013). Soft systems methodology. In *Encyclopedia of operations research and management science* (pp. 1430-1436). Springer.

Chen, C.-h., Härdle, W., Unwin, A., Kerman, J., Gelman, A., Zheng, T., & Ding, Y. (2008). Visualization in Bayesian data analysis. *Handbook of data visualization*, 709-724.

Christensen, R., Johnson, W., Branscum, A., & Hanson, T. E. (2011). *Bayesian ideas and data analysis: an introduction for scientists and statisticians*. CRC press.

Cianciolo, A. T., Matthew, C., Sternberg, R. J., & Wagner, R. K. (2006). Tacit knowledge, practical intelligence, and expertise. *The Cambridge handbook of expertise and expert performance*, 613-632.

Collazo, R., Görgen, C., & Smith, J. Q. (2018). *Chain Event Graphs*.

Consonni, G., Fouskakis, D., Liseo, B., & Ntzoufras, I. (2018). Prior Distributions for Objective Bayesian Analysis. *Bayesian Analysis*, *13*(2), 627 - 679.

Constable, A. J., French, S., Karoblyte, V., & Viner, D. (2022). Decision-Making for Managing Climate-Related Risks: Unpacking the Decision Process to Avoid "Trial-and-Error" Responses. *Frontiers in Climate*, 84.

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review*, *70*(2), 161-189.
De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'institut Henri Poincaré*, *7*(1), 1-68.
De Finetti, B. (1974). *Theory of Probability* (Vol. 1). John Wiley and Sons.
De Finetti, B. (1975). *Theory of Probability* (Vol. 2). John Wiley and Sons.
de Neys, W. (2021). On dual and single process models of thinking. *Perspectives on Psychological Science*(In Press).
DeGroot, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill.
Dias, L., Morton, A., & Quigley, J. (Eds.). (2018). *Elicitation of Preferences and Uncertainty: Processes and Procedures*. Springer.
Draper, D. (1995). Assessment and propagation of model uncertainty (with discussion). *Journal of the Royal Statistical Society*, *B57*(1), 45-97.
Eden, C. (1988). Cognitive mapping: a review. *European Journal of Operational Research*, *36*, 1-13.
Eden, C., & Radford, J. (Eds.). (1990). *Tackling Strategic Problems: the Role of Group Decision Support*. Sage.
Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*, 380-417.
Edwards, W., Miles, R. F., & Von Winterfeldt, D. (Eds.). (2007). *Advances in Decision Analysis: from Foundations to Applications*. Cambridge University Press.
Evans, J. S. B., & Stanovich, K. E. J. P. o. p. s. (2013). Dual-process theories of higher cognition: Advancing the debate. *8*(3), 223-241.
Fine, T. L. (1973). *Theories of Probability*. Academic Press.
Franco, L. A., & Montibeller, G. (2010). Facilitated modelling in operational research. *European Journal of Operational Research*, *205*(3), 489-500.
French, S. (1978). A Bayesian three-stage model in crystallography. *Acta Crystalographica*, *A34*, 728-738.
French, S. (1986). *Decision Theory: an Introduction to the Mathematics of Rationality*. Ellis Horwood.
French, S. (1995). Uncertainty and imprecision: modelling and analysis. *Journal of the Operational Research Society*, *46*(1), 70-79.
French, S. (2003). Modelling, making inferences and making decisions: the roles of sensitivity analysis. *TOP*, *11*(2), 229-252.
French, S. (2013). Cynefin, Statistics and Decision Analysis. *Journal of the Operational Research Society*, *64*(4), 547-561. https://doi.org/doi:10.1057/jors.2012.23
French, S. (2015). Cynefin: Uncertainty, Small Worlds and Scenarios. *Journal of the Operational Research Society*, *66*(10), 1635-1645.
French, S. (2021). From Soft to Hard Elicitation. *Journal of the Operational Research Society*, *73*(6), 1181-1197.
French, S. (2022). Axiomatising the Bayesian Paradigm in Parallel Small Worlds. *Operations Research*, *70*(3), 1342-1358.
French, S. (2023). Reflections on 50 years of MCDM: Issues and Future Research Needs. *European Journal on Decision Processes*, *11*(Published online), 100030. https://doi.org/https://doi.org/10.1016/j.ejdp.2023.100030
French, S. (2024). Whose Judgement? Reflections on Elicitation in Bayesian Analysis (with discussion). *Decision Analysis*, *21*(3), 143-164.
French, S., & Argyris, N. (2018). Decision Analysis and Political Processes. *Decision Analysis*, *15*(4), 208-222.
French, S., Haywood, S., Oughton, D. H., & Turcanu, C. (2020). Different types of uncertainty in nuclear emergency management. *Radioprotection*, *55*, S175-S180.
French, S., Maule, A. J., & Papamichail, K. N. (2009). *Decision Behaviour, Analysis and Support*. Cambridge University Press.
French, S., & Rios Insua, D. (2000). *Statistical Decision Theory*. Arnold.
French, S., & Wilson, K. S. (1978). On the treatment of negative intensity observations. *Acta Crystalographica*, *A34*, 517-525.
Freund, R. J., Vail, R. W., & Clunies-Ross, C. (1961). Residual analysis. *Journal of the American Statistical Association*, *56*(293), 98-104.
Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in Bayesian workflow. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *182*(2), 389-402.
Gelman, A. (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review*, *71*(2), 369-382.
Gelman, A. (2004). Exploratory Data Analysis for Complex Models. *Journal of Computational and Graphical Statistics*, *13*(4), 755-779.

Gelman, A. (2018). Ethics in statistical practice and communication: Five recommendations. *Significance*, *15*(5), 40-43.

Gelman, A., & Hennig, C. (2017). Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *180*(4), 967-1033.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y.,…Modrák, M. (2020). Bayesian workflow. *arXiv preprint arXiv:2011.01808*.

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2009). *Research Methods in Applied Settings: An Integrated Approach to Design and Analysis*. Routledge - Taylor and Francis.

Goldstein, M., & Rougier, J. C. (2009). Reified Bayesian Modelling and Inference for Physical Systems (with Discussion). *Journal of Statistical Planning and Inference*, *139*, 1221-1256.

Good, I. J. (1983). The philosophy of exploratory data analysis. *Philosophy of Science*, *50*(2), 283-295.

Granger Morgan, M., & Henrion, M. (1990). *Uncertainty: a guide to dealing with Uncertainty in Qualitative Risk and Policy Analysis*. Cambridge University Press.

Gregory, R. S., Failing, L., Harstone, M., Long, G., McDaniels, T., & Ohlson, D. (2013). *Structured Decision Making: A Practical Guide to Environmental Management Choices*. Wiley-Blackwell.

Gui, Y., Jin, Y., & Ren, Z. (2024). Conformal alignment: Knowing when to trust foundation models with guarantees. *arXiv preprint arXiv:2405.10301*.

Hand, D., Mannila, H., & Smyth, P. (2001). *Data Mining*. MIT Press.

Hand, D. J. (2020). *Dark data*. Princeton University Press.

Hanea, A., Nane, G. F., Bedford, T., & French, S. (2021). *Expert Judgement in Risk and Decision Analysis*. Springer.

Hennig, P., Osborne, M. A., & Girolami, M. (2015). Probabilistic numerics and uncertainty in computations. Proc. R. Soc. A,

Hubbard, R., Haig, B. D., & Parsa, R. A. (2019). The limited role of formal statistical inference in scientific inference. *The American Statistician*, *73*(sup1), 91-98.

Jeffreys, H. (1961). *Theory of Probability* (3 ed.). Oxford University Press.

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, *29*, 36-52.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin, Allen Lane.

Kahneman, D., & Tversky, A. (1974). Judgement under uncertainty: heuristics and biases. *Science*, *185*(4157), 1124-1131.

Katsikopoulos, K. V. (2011). Psychological heuristics for making inferences: Definition, performance, and the emerging theory and practice. *Decision analysis*, *8*(1), 10-29.

Keeney, R. L. (1992a). On the foundations of prescriptive decision analysis. In W. Edwards (Ed.), *Utility theories: Measurements and applications* (pp. 57-72). Kluwer Academic Publishers.

Keeney, R. L. (1992b). *Value-Focused Thinking: a Path to Creative Decision Making*. Harvard University Press.

Keeney, R. L. (2012). Value-focused brainstorming. *Decision Analysis*, *9*(4), 303-313.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.

Klosgen, W., & Zytkow, J. M. (Eds.). (2002). *Handbook of Datamining and Knowledge Discovery*. Oxford University Press.

Knight, F. H. (1921). *Risk, Uncertainty and Profit*. Hart, Schaffner & Marx; Houghton Mifflin Company.

Kuhn, T. S. (1961). The function of measurement in modern physical science. *Isis*, *52*(2), 161-193.

Lei, L., & Candès, E. J. (2021). Conformal inference of counterfactuals and individual treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *83*(5), 911-938.

Lindley, D. V. (1992). Is our View of Bayesian Statistics too Narrow? (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 1-15). Oxford University Press.

Lord Phillips. (2000). *The Phillips Report on the BSE Crisis*.

Marchau, V., Walker, W. E., Bloemen, P., & Popper, S. (Eds.). (2019). *Decision Making under Deep Uncertainty*. Springer.

Martin, J., De Adana, D. D. R., & Asuero, A. G. (2017). Fitting models to data: residual analysis, a primer. In J. P. Hessling (Ed.), *Uncertainty quantification and model calibration* (Vol. 133). IntechOpen.

Marttunen, M., Lienert, J., & Belton, V. (2017). Structuring problems for Multi-Criteria Decision Analysis in practice: A literature review of method combinations. *European Journal of Operational Research*, *263*(1), 1-17. https://doi.org/https://doi.org/10.1016/j.ejor.2017.04.041

Miller, G. A. (1956). The magic number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81-97.

Montibeller, G., & Winterfeldt, D. (2015). Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis*, *35*(7), 1230-1251.

Morton, A., & Fasolo, B. (2009). Behavioural decision theory for multi-criteria decision analysis: a guided tour 60(2): 268-275. *Journal of the Operational Research Society*, *60*(2), 268-275.

National Academies of Sciences, E., and Medicine. (2019). *Reproducibility and Replicability in Science*. The National Academies Press. https://doi.org/doi:10.17226/25303

Nonaka, I. (1991). The knowledge creating company. *Harvard Business Review*, *6*(8), 96 - 104.

O'Hagan, A. (1992). Some Bayesian Numerical Analysis (with Discussion). In J. Bernardo, J. Berger, A. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4* (pp. 345-363). Oxford University Press.

Oatley, S. J., & French, S. (1982). A profile-fitting method for the analysis of diffractometer data. *Acta Crystalographica*, *A38*, 537-549.

Oliver, R. M., & Smith, J. Q. (Eds.). (1990). *Influence Diagrams, Belief Nets and Decision Analysis*. John Wiley and Sons.

O'Hagan, A. (2019). Expert knowledge elicitation: subjective but scientific. *The American Statistician*, *73*(sup1), 69-81.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Phillips, L. D. (1984). A theory of requisite decision models. *Acta Psychologica*, *56*(1-3), 29-48.

Phillips, L. D. (2007). Decision conferencing. In W. Edwards, R. F. Miles, & D. von Winterfeldt (Eds.), *Advances in Decision Analysis: from Foundations to Applications* (pp. 375-399). Cambridge University Press.

Phillips, L. D. (2011). Group dynamics processes for improved decision making. In J. J. Cochran, L. A. Cox Jr., P. Keskinocak, J. P. Kharoufeh, & J. Cole Smith (Eds.), *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley and Sons.

Phillips, L. D., & Phillips, M. C. (1993). Facilitated Work Groups - Theory and Practice. *Journal of the Operational Research Society*, *44*(6), 533-549.

Polyani, M. (1962). *Personal Knowledge*. Anchor Day Books.

Ramsey, F. P. (1926). Truth and Probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays*. Harcourt, Brace and Co.

Reagan-Cirincione, P. (1994). Combining group facilitation, decision modelling and information technology to improve the accuracy of group judgement Improving the accuracy of group judgment: a process intervention combining group facilitation, social judgment analysis, and information technology. *Organisational Behaviour and Human Decision Process*, *58*, 246-270.

Riesch, H. (2013). Levels of uncertainty. In *Essentials of risk theory* (pp. 29-56). Springer.

Rios Insua, D., & Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer-Verlag.

Roberts, F. S. (1979). *Measurement Theory*. Academic Press.

Rogers, S., & Girolami, M. (2015). *A first course in machine learning*. CRC Press.

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, *14*(11), e12633.

Rosenhead, J., & Mingers, J. (Eds.). (2001). *Rational Analysis for a Problematic World Revisited*. John Wiley and Sons.

Savage, L. J. (1972). *The Foundations of Statistics* (2nd ed.). Dover.

Schwab, S., Janiaud, P., Dayan, M., Amrhein, V., Panczak, R., Palagi, P. M.,…Senn, S. (2022). Ten simple rules for good research practice. In (Vol. 18, pp. 14): Public Library of Science San Francisco, CA USA.

Shafer, G. (1986). Savage revisited. *Statistical science*, 463-485.

Shaw, D., Franco, A., & Westcombe, M. (2006). Special Issue: Problem Structuring Methods I. In *Journal of the Operational Research Society* (Vol. 57, pp. 757-878).

Shaw, D., Franco, A., & Westcombe, M. (2007). Special Issue: Problem Structuring Methods II. In *Journal of the Operational Research Society* (Vol. 58, pp. 545- 682).

Shleifer, A. (2012). Psychologists at the Gate: A Review of Daniel Kahneman's" Thinking, Fast and Slow". *Journal of Economic Literature*, *50*(4), 1080-1091.

Silvey, S. D. (1975). *Statistical inference* (2nd ed.). CRC Press.

Smith, C. M., & Shaw, D. (2019). The characteristics of problem structuring methods: A literature review. *European Journal of Operational Research*, *274*(2), 403-416.

Snowden, D. (2002). Complex acts of knowing - paradox and descriptive self-awareness. *Journal of Knowledge Management*, *6*(2), 100-111.

Spiegelhalter, D. (2019). *The art of statistics: How to learn from data*. Basic Books.

Stark, P. B., & Saltelli, A. (2018). Cargo-cult statistics and scientific crisis. *Significance*, *15*(4), 40-43.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-463.

Walker, W. E., Lempert, R. J., & Kwakkel, J. H. (2013). Deep uncertainty. In S. Gass & M. C. Fu (Eds.), *Encyclopaedia of Operations Research and Management Science* (pp. 395-402). Springer.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, *70*(2), 129-133.

Wilkerson, R. L., & Smith, J. Q. (2020). Customised Structural Elicitation. In A. Hanea, G. F. Nane, T. Bedford, & S. French (Eds.), *Expert Judgement in Risk and Decision Analysis*. Spinger.

Zhang, X., Simpson, T., Frecker, M., & Lesieutre, G. (2010). Supporting knowledge exploration and discovery in multi-dimensional data with interactive multiscale visualisation. *Journal of Engineering Design*, *23*(1), 23-47.