

When Did Bayesian Inference Become “Bayesian”?

Stephen E. Fienberg*

Abstract. While Bayes’ theorem has a 250-year history, and the method of inverse probability that flowed from it dominated statistical thinking into the twentieth century, the adjective “Bayesian” was not part of the statistical lexicon until relatively recently. This paper provides an overview of key Bayesian developments, beginning with Bayes’ posthumously published 1763 paper and continuing up through approximately 1970, including the period of time when “Bayesian” emerged as the label of choice for those who advocated Bayesian methods.

Keywords: Bayes’ Theorem; Classical statistical methods; Frequentist methods; Inverse probability; Neo-Bayesian revival; Stigler’s Law of Eponymy; Subjective probability.

1 Introduction

What’s in a name? It all depends, especially on the nature of the entity being named, but when it comes to statistical methods, names matter a lot.

Whether the name is eponymous (as in Pearson’s chi-square statistic¹, Student’s t -test, Hotelling’s T^2 statistic, the Box-Cox transformation, the Rasch model, and the Kaplan-Meier statistic) or “generic” (as in correlation coefficient or p -value) or even whimsical (as in the jackknife² or the bootstrap³), names in the statistical literature often signal the adoption of new statistical ideas or shifts in the acceptability of statistical methods and approaches.⁴ Today statisticians speak and write about Bayesian statistics and frequentist or classical statistical methods, and there is even a journal of *Bayesian Analysis*, but few appear to know where the descriptors “Bayesian” and “frequentist” came from or how they arose in the history of their field.

This paper is about the adjective “Bayesian”⁵ and its adoption by the statistical community to describe a set of inferential methods based directly on the use of *Bayes’ Theorem*, which is now thought of by many as an elementary result in probability

*Department of Statistics, CyLab, and Center for Automated Learning and Discovery, Carnegie Mellon University, Pittsburgh, PA, <http://www.stat.cmu.edu/~fienberg>

¹Many authors take great pride in having an eponymous method in their lifetime, but this was insufficient for Karl Pearson. See Porter (127, Chapter 1).

²Named by John Tukey after the boy scout’s trusty knife.

³Coined by Bradley Efron by reference to the tales of the adventures of Baron Munchausen, who rescued himself from drowning in a lake by picking himself up by his own bootstraps.

⁴Kasner and Newman (90) noted: “We can get along without new names until, as we advance in science, we acquire new ideas and new forms.”

⁵Bayesian is also now used as a noun, as in *a Bayesian*, i.e., a person who thinks it makes sense to treat observables as random variables and to assign probability distributions to them. Such usage followed the adoption of “Bayesian” as an adjective.

theory and where probabilities are typically taken to be subjective or logical. The paper is not about the question “Who discovered Bayes’ Theorem?,” an issue related to the emergence of statistical thinking some 250 years ago, and one evoking references to Stigler’s Law of Eponymy⁶. Indeed, Stigler (158) has addressed this question himself, as have others such as Dale (36) and Bellhouse (16). Rather, my focus is on the emergence of the adjective “Bayesian” as associated with inferential methods in the middle of the twentieth century to describe what has been referred to up to that point as the method of *inverse probability*. Why did the change occur? To whom should the term and its usage be attributed? What was the impact of the activities surrounding the adoption of the adjective “Bayesian”? Why do many statisticians now refer to themselves as Bayesian?⁷ These are some of the questions I plan to address. Aldrich (2) covers some closely related territory but with a different focus and perspective.

The task of investigating the usage of names was once quite daunting, but the recently-developed fully-searchable electronic archives such as JSTOR have made the task of exploring word usage far simpler than it once was, at least for English language statistical literature. Nonetheless, usage occurs in books and oral presentations, as well as in informal professional communications, and a search for terms like “Bayesian” in only electronic journal archives, while informative, would be unlikely to answer the key question I want to address. Pieces of the beginning of my story have been chronicled in different forms in the histories by Dale (36), Hald (82), Porter (126), and Stigler (159; 161), but these otherwise wonderful sources do not answer the key question that gives its name to the title of this paper. I began asking these questions several years ago (e.g., see Fienberg (58)) as a result of conversations with a social historian collaborator about the origin of methods for census adjustment. I had the good fortune to be able to confer with several distinguished statisticians who participated in some key events which I will describe, and their recollections (e.g., see Lindley (106)) have contributed to the answers I provide here.

In the next two sections, I trace some of the history of Bayesian ideas, from the time of Bayes and Laplace through to the twentieth century. In Section 4, I turn to the developments of the first half of the twentieth century, both Bayesian and otherwise, since they set the stage for the neo-Bayesian revival. In section 5, I explain what is known about the use of adjective “Bayesian,” and I outline the dimensions of Bayesian creativity that surrounded its emergence during a period I call the neo-Bayesian revival. In the final sections of the paper, I focus briefly on some of the sequelae to the neo-

⁶Stigler’s Law in its simplest form states that “[n]o scientific discovery is named after its original discoverer.” For those who have not previously seen references to this law, it is worth noting that Stigler proposed it in the spirit of a self-proving theorem. (156)

⁷A reviewer of an earlier version of this article observed that this is a decidedly English-language, and perhaps uninteresting, question. Making adjectives out of nouns was perhaps more common in German than in English in the late nineteenth and early twentieth centuries, when the adjective “Bayessche” was used; however, “der Bayesschen Satz” in Kolmogorov’s *Grundbegriffe* (93, p. 36) is just a theorem about events and not Bayesian in the sense that I use it here (c.f. Aldrich, in *Earliest Known Uses of Some of the Words of Mathematics*, <http://members.aol.com/jeff570/b.html>). Moreover, while there may not be much apparent difference between “Bayes rule” and “Bayesian rule,” the adoption of the adjective in English occurred during a remarkable period of intellectual ferment and marked the rise of the modern Bayesian movement in a predominantly English-language literature.

Bayesian revival of the 1950s and especially during the 1960s. Today, Bayesian methods are integrated into both the fabric of statistical thinking within the field of statistics and the methodology used in a broad array of applications. The ubiquity of Bayesian statistics is illustrated by the name of the *International Society for Bayesian Analysis*, its growing membership, and its new on-line journal. But one can also see the broad influence of Bayesian thinking by a quick scan of major journals of not only statistics but also computer science and bioinformatics, economics, medicine, and even physics, to name specific fields.

This paper is far from an exhaustive treatment of the subject, for that would have taken a book. Rather, I have chosen to cite a few key contributions as part of the historical development, especially as they relate to the theme of the adjective “Bayesian.” I cite many (but not all) relevant books and a small fraction of the papers that were part of the development of Bayesian inference.

2 Bayes’ Theorem

My story begins, of course, with the Reverend Thomas Bayes,⁸ a nonconformist English minister whose 1763 posthumously published paper, “An Essay Towards Solving a Problem in the Doctrine of Chances,” (14) contains what is arguably the first detailed description of the theorem from elementary probability theory now associated with his name. Bayes’ paper, which was submitted for publication by Richard Price, is remarkably opaque in its notation and details, and the absence of integral signs makes for difficult reading to those of us who have come to expect them.

The Essay considers the problem, “*Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.” [p. 376] Writing in an unpublished 1960 reading note, L.J. Savage (141) observed: “The problem is of the kind we now associate with Bayes’s name, but it is confined from the outset to the special problem of drawing the Bayesian inference, not about an arbitrary sort of parameter, but about a ‘degree of probability’ only.” This statement actually provides us with the first clue to the title of this article; clearly in 1960, Savage was using the term “Bayesian” as we do today. And he notes what others have subsequently: that Bayes did not actually give us a statement of Bayes’ Theorem, either in its discrete form,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}, \quad (1)$$

(this came later in Laplace (96)), or in its continuous form with integration, although he solved a special case of the latter.

In current statistical language, Bayes’ paper introduces a uniform prior distribution on the binomial parameter,⁹ θ , reasoning by analogy with a “billiard table” and drawing

⁸For biographical material on Bayes see Bellhouse (16) and Dale (37).

⁹Of course Bayes didn’t use the term *parameter*—David and Edwards (41) trace the introduction

on the form of the marginal distribution of the binomial random variable, and not on the principle of “insufficient reason,” as many others have claimed.¹⁰ An appendix to the paper, written by Price, also deals with the related problem of predicting the result of a new observation. Bellhouse (15) suggests that Price contributed more to Bayes’ essay than others have hitherto suggested, perhaps even piecing it together from a number of “initially seemingly unrelated results” to form a coherent whole. Both Bellhouse and Stigler have also suggested that the famous portrait of Bayes that adorns a myriad of publications about him and all things Bayesian may actually not be Bayes at all (see (122))!

Stigler (158) has also explored the question of whether Bayes did indeed discover Bayes’ Theorem, and points to paragraphs in a 1749 book by David Hartley that give a concise description of the inverse result and attribute it to a friend, whom Stigler infers was likely Nicholas Saunderson, but if so he apparently did not publish the details. Hald (82) argues that the friend might actually have been Bayes and that the result was developed before the publication of Hartley’s book. Dale (37) reproduces an undated notebook of Bayes with a passage which he suggests comes from about 1746-1749, thus supporting the thesis that Bayes might have been the unnamed friend of Hartley.

Would we call Bayes a “Bayesian” today? Stigler (157) argues that Bayes intended his results in a rather more limited way than would modern Bayesians. But Stigler also notes that Bayes’ definition of probability is subjective, and a 2003 version of the entry on Thomas Bayes in the online *Wikipedia Encyclopedia* entry on suggested that we interpret it in terms of expected utility (had Bayes only understood the concept!), and thus that Bayes’ result would make sense only to the extent to which one can bet on its observable consequences. But perhaps it would be fairer to Bayes to see what ideas unfolded in the two centuries following his death before probing more deeply into how he viewed the matter! Nonetheless this commentary on Bayes raises anew the issue of why we call Bayesian methods “Bayesian.”

3 Inverse Probability From Bayes to the Twentieth Century

Whether or not Bayes actually discovered Bayes’ Theorem, it seems clear that his work preceded that of Pierre Simon Laplace, the eighteenth century French scientist who, in his 1774 paper, “Mémoire sur la Probabilité des Causes par les Évènements,” gave a much more elaborate version of the inference problem for the unknown binomial parameter in relatively modern language and notation. Laplace also articulated, more clearly than Bayes, his argument for the choice of a uniform prior distribution, arguing that the posterior distribution of the parameter θ should be proportional to what we

of the term to a 1903 book by Kapteyn (89). Then in 1922 Fisher (62) reintroduced the term and the label took hold. For modern Bayesians, parameters are simply random variables and so it is natural to put distributions on them.

¹⁰See Stigler (157), who draws attention to this marginal distribution argument in the “Scholium” that follows the key proposition in Bayes’ paper.

now call the likelihood of the data, i.e.,

$$f(\theta|x_1, x_2, \dots, x_n) \propto f(x_1, x_2, \dots, x_n|\theta). \quad (2)$$

We now understand that this implies that the prior distribution for θ is uniform, although in general, of course, the prior may not exist. The paper also contains other major statistical innovations, such as picking an estimate that minimizes the posterior loss. For further details, see Stigler (160).

Laplace refined and developed the “Principle” he introduced in 1774 in papers published in 1781 and 1786, and it took on varying forms such as the “indifference principle” or what we now refer to as “Laplace’s Rule of Succession” (for obtaining the probability of new events on the basis of past observations). But the original 1774 memoir had far-reaching influence on the adoption of Bayesian ideas in the mathematical world, influence that was unmatched by Bayes’ paper, to which it did not refer.¹¹ Ultimately, Laplace and others recognized Bayes’ prior claim (e.g., see Condorcet’s introduction to Laplace’s 1781 paper, and the final section of Laplace’s 1812 monograph (97), *Théorie Analytique des Probabilités*, and the related discussion in Dale (36), p. 249). Condorcet (33) used a form probabilistic reasoning attributed to Bayes and Laplace in his famous work on majority voting, as did his student Lacroix (95) in his probability text.

Laplace’s introduction of the notion of “indifference” as an argument in specifying a prior distribution was first in a long line of efforts to discover the statistical holy grail: prior distributions reflecting ignorance. The search for the holy grail continues today under a variety of names, including objective Bayes, and in response to every method that is put forward we see papers about logical and statistical inconsistencies or improprieties.

Another surprise to some is that Laplace’s 1774 paper does not use the term *inverse probability*; the phrase came into use later. For example, De Morgan (47) wrote about the method of inverse probability and attributed its general form to Laplace’s 1812 book (97)—“inverse” because it involves inferring backwards from the data to the parameter or from effects to causes. The term continued in use until the mid-twentieth century. It was only later that others argued that Laplace had not noticed that inverting the conditioning changed the meaning of probability (c.f. the narrative in Howie (85)).

Laplace’s 1812 formulation of the large sample normal probability approximation for the binomial parameter with a uniform (beta) prior was generalized by I.J. Bienaymé (17) to linear functions of multinomial parameters, again with a uniform prior on the multinomial. In essence, he was working with a special case of the Dirichlet prior. Heyde and Seneta (83, pp. 97–103) give a detailed description of Bienaymé’s formulation and approach in modern notation and they note that a rigorous proof of the result was provided in the twentieth century by von Mises.¹²

¹¹See Stigler (160) for a discussion of why Laplace was unaware of Bayes’ paper when he wrote this 1774 paper.

¹²Results similar to this one by Bienaymé are used today to justify frequentist interpretations of Bayesian interval estimates. It is in this sense that Bienaymé was clearly before his time, as the title of (83) suggests.

According to Daston (38), “[B]etween 1837 and 1843 at least six authors—Siméon-Denis Poisson, Bernard Bolzano, Robert Leslie Ellis, Jacob Friedrich Fries, John Stuart Mill, and [Antoine Augustine] Cournot—approaching the topic as mathematicians, writing in French, German, and English, and apparently working independently made similar distinctions between the probabilities of things and the probabilities of our beliefs about things.” This is the distinction between what we now call objective and subjective meanings of probability. Cournot (35), for example, sharply criticized Condorcet’s and Laplace’s approach to the probabilities of causes (see Daston (38) for a detailed discussion of what all of these authors meant by these terms). This focus on the meaning of probability quickly spread to other countries as well. The links between France and Russia were especially strong as Seneta (150) notes and, while the work in Russia influenced the formal development of probability in the twentieth century in profound ways,¹³ it seems to have had limited impact at the time on the development of subjective probability and related statistical methods. This is not to suggest that the developments in probability were devoid of statistical content. For example, Bienaymé in 1853 and Chebyshev in 1867, in proofs of what is most commonly known as the Chebyshev inequality, independently set out what could be thought of as the *Method of Moments* (e.g., see Heyde and Seneta (83, pp. 121–124)). But both authors were primarily interested in forms of the Law of Large Numbers and not problems of statistical inference per se. It was Karl Pearson, decades later, who set forth the method in the context of inferential problems and actual data. The connection of these developments to inverse probability is, however, somewhat muddy.

Mathematicians and philosophers continued to debate the meaning of probability throughout the remainder of the nineteenth century. For example, Boole, Venn, Chrystal, and others argued that the inverse method was incompatible with objective probabilities, and they also critiqued the inverse method because of its use of indifference prior distributions. But when it came to the practical application of statistical ideas for inferential purposes, inverse probability ruled the day (c.f. Zabell (181; 182)).¹⁴ This was largely because no one came forward with a systematic and implementable “frequentist” alternative. For example, Fechner (55) studied frequency distributions and introduced the term *Kollektivmasslehre*—collective—later used by von Mises in his frequentist approach. But, as Sheynin (153) notes, his description was vague and general in nature.

Thus, in retrospect, it shouldn’t be surprising to see inverse probability as the method of choice of the great English statisticians of the turn of the century, such as Edgeworth and Pearson. For example, Edgeworth (49) gave one of the earliest derivations of what we now know as Student’s *t*-distribution, the posterior distribution of the mean μ of a normal distribution given uniform prior distributions on μ and $h = \sigma^{-1}$, calling it the “subexponential” distribution. This is another instance of Stigler’s Law, since Gosset, publishing under the pseudonym *Student*, gave his version of the derivation

¹³For example, see the discussion about the evolution of ideas in probability in Shafer and Vovk (151).

¹⁴For example, according to Sheynin (152), Markov used a variation on an argument of Laplace to demonstrate an early “robustness” argument with regard to prior distributions in the binomial setting.

in 1908 (163), without reference to Edgeworth (c.f. Stigler (155)).

Gosset’s exposition does not directly mention inverse probability, but it is reasonably clear to anyone who reads his subsequent paper on the correlation coefficient, published the same year, that this was the implicit reasoning (see Fienberg and Lazar (60)). Of course, today we also know that the result was derived even earlier, by Lüroth in 1876, who also used inverse probability methods (see Pfanzagl and Sheynin (125)).

Gosset worked at Guinness Brewery in Dublin, but did the work on the t -distribution and the correlation coefficient in Karl Pearson’s Statistical Laboratory at University College London. Beginning in his *Grammar of Science* (123), Pearson adopted the Laplace version of inverse probability, but he also argued for the role of experience in determining the *a priori* probabilities. In Pearson (124), he again wrote about assuming “the truth of Bayes’ Theorem,” and went on to discuss the importance of past experience. This is an approach which, when implemented decades later, came to be known as empirical Bayes. But, as Jeffreys noted much later on when describing Pearson’s position,

[t]he anomalous feature of his work is that though he always maintained the principle of inverse probability . . . he seldom used it in actual applications, and usually presented his results in a form that appears to identify a probability with a frequency. (87, p. 383)

It should come as no surprise to us therefore that those like Gosset, on whom Pearson exerted such a strong influence, would use inverse probability implicitly or explicitly. Yet later, Gosset would gladly assume the mantle of frequentist methods advocated by R. A. Fisher, although he, like most other statisticians, found issues on which he and Fisher disagreed quite strongly.¹⁵

4 From Inverse Probability to Frequentism and the Rise of Subjective Probability

4.1 Frequentist Alternatives to Inverse Probability

At the time that Ronald Alymer Fisher began his studies of statistics at Cambridge in 1909, inverse probability was an integral part of the subject he learned (c.f. Edwards (50)). Frequentist and other non-Bayesian ideas were clearly “in the air,” but it is difficult to know to what extent Fisher was aware of them. For example, as a student he might have been led to read papers and books dealing with such alternatives by his teacher, F.J.M. Stratton, but he rarely cited relevant precursors or alternatives to his own work (see Aldrich’s (1) account of what Fisher studied). Over a period of 10

¹⁵Fienberg and Lazar (60) provide support for this interpretation of Gosset as using inverse probability. A reviewer of an earlier version of this paper has argued that Gosset was simply not clear about these matters and that his later acceptance of Fisher’s interpretation of inference was not foreordained. But such an argument misses the fact that there was no comprehensive statistical alternative to inverse probability at the time Gosset wrote his 1908 papers and thus we should expect to see the role of inverse probability in his work.

years, from 1912 to 1922, Fisher moved away from the inverse methods and towards his own approach to inference he called the “likelihood,” a concept he claimed was distinct from probability. But Fisher’s progression in this regard was slow. Stigler (162) has pointed out that, in an unpublished manuscript dating from 1916, Fisher didn’t distinguish between likelihood and inverse probability with a flat prior, even though when he later made the distinction he claimed to have understood it at this time. But within six years Fisher’s thinking had a broad new perspective, and his 1922 paper (62) was to revolutionize modern statistical thinking. In it, he not only introduced likelihood and its role in producing maximum likelihood estimates, but he also gave us the statistical notions of “sufficiency” and “efficiency” and used the label “parameter,” which was to become the object of scientific inquiry (c.f. Stigler (162)). Later he moved even further and developed his own approach to inverse reasoning which he dubbed the “fiducial” method (64), and went so far as to suggest that Gosset’s interpretation of his t -distribution result was in the fiducial spirit. Fisher also gave statistics the formal methodology of tests of significance, especially through his 1925 book, *Statistical Methods for Research Workers* (63).

Fisher’s work had a profound influence on two other young statisticians working in Pearson’s laboratory at University College London: Jerzy Neyman, a Polish statistician whose early work focused on experiments and sample surveys, and Egon Pearson, Karl Pearson’s son. They found Fisher’s ideas on significance tests lacking in mathematical detail and, together, they set out to extend and “complete” what he had done. In the process, they developed the methods of hypothesis testing and confidence intervals that were to revolutionize both the theory and the application of statistics. Although Fisher often disagreed with them caustically and vigorously both orally and in print (e.g., see Fienberg and Tanur (61)), some amalgam of their approaches—referred to later as “frequentist” methods—soon supplanted inverse probability. The mathematical foundation for these methods emerged from Russia and the axiomatic work of Kolmogorov (93). The methods quickly spread to diverse areas of application, for good and for bad. Gigerenzer et al. (69) describe the impact of these frequentist methods especially in psychology. I would be remiss if I didn’t also mention the work of Richard von Mises (168) on frequentist justifications for probability, but his work had remarkably little influence on the course of the development of statistical methods and theory despite its seeming importance to philosophers. In his 1941 paper on the foundations of probability and statistics, however, von Mises (169) uses a Bayesian argument to critique the Neyman method of confidence intervals. While he suggested that what one really wanted was the posterior distribution, the estimation challenge was saying something about it “without having information about the prior probability,” a theme he elaborated on in von Mises (170).

The adjective “frequentist,” like the adjective “Bayesian,” is now widely used, but it was uncommon in the early days of Fisher, Neyman, and Pearson. A search of JSTOR shows that its earliest use in the main English language statistical journals occurred in a 1936 article by Nagel (119); his focus was largely on the frequency interpretation of probability. The adjective appeared again sporadically in philosophy and statistics journals for the next 20 years, but the use of “frequentist” to describe statistical methods

gained currency in the 1950s only after “Bayesian” came into common usage, and then it was used by Bayesians to describe non-Bayesian methods.¹⁶

Abraham Wald used Bayesian ideas and the name “Bayes” throughout his development of statistical decision theory, albeit in a decidedly non-Bayesian or frequentist form. Wald (172) considered “hypothetical a priori distributions” of the parameter of interest, and in 1947 (173) he derived the “Bayes solution,” but then evaluated its performance with respect to the sample space. The basic idea in all of this work was that the use of Bayesian tools led to good frequentist properties. Later literature renamed other quantities in Wald’s 1939 paper using “Bayes” as a descriptor. Writing in 1953 as a discussant to a paper by Dennis Lindley (100), George Barnard was to remark that “Wald helped to rescue Bayes’ theorem from the obscurity into which it had been driven in recent years.”

4.2 The Rise of Subjective Probability

Of course, inverse probability ideas did not die with the rise of likelihood methods and tests of hypotheses. John Maynard Keynes (91) described the view of subjective probability in the early 1920s in his *Treatise on Probability*, but in doing so he allowed for the possibility that degree of belief might not be numerically measurable. While Keynes’ bibliography ran some 25 pages and made clear that he was drawing on a broad literature, in other senses it signaled a major break from past, in large part because of the new literature it triggered.¹⁷ In his review of Keynes’ book, Émile Borel (21) clearly identified the issues associated with the meaning of probability and articulated what Savage was to refer to as “the earliest account of the modern concept of personal probability known to me.”¹⁸ Five years later, Frank Ramsey (132) critiqued

¹⁶The term “classical” statistics, used to describe frequentist as opposed to Bayesian methods, came later, and perhaps it owes itself to Neyman’s 1937 paper (121), “Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability.” Neyman used the adjective “classical” to distinguish his approach from Jeffreys’ and suggested that his approach was rooted in the traditional approach to probability based on equally likely cases or the more formal versions developed in the twentieth century by the French and Russian statisticians Borel, Fréchet, Kolmogorov, and Lévy. In the process he in some senses usurped the label and applied it to his approach with Egon Pearson. This usage of “classical” in an inference context is an oxymoron, since if anything should be called classical it should be methods descended from inverse probability. Wilks (178) used the phrase “classical probability” in describing sampling distributions in the context of confidence interval calculations the previous year in his discussion of Nagel (119), so again the use of the label applied to inference methods was “in the air.” Various authors have of course used the label “classical” to refer to methods from the past. But what is clear from a search of JSTOR is that “classical inference,” “classical statistical inference,” and “classical statistics” came into vogue in the 1950s and 1960s, especially by Bayesians to describe non-Bayesian frequentist methodology associated with the Neyman-Pearson school. Barnett (13) has a chapter in the first edition of his comparative inference book on “classical inference.”

¹⁷A reviewer criticized an earlier version of this paper as being Anglo-centric. But the literature that we now label as Bayesian, with a few notable exceptions, became dominated by English-speaking authors whose sources typically did not include the diverse array of references which Keynes cited. This is not to argue that Keynes and those who followed him were not influenced by the French school, but rather that the French literature stopped far short of providing the foundation on which modern Bayesian thinking now rests.

¹⁸See the comment in the bibliographic supplement to the paperback edition of Savage (140). A referee has pointed to similar interpretations in Paul Lévy’s 1925 book on probability (98), although

Keynes’s axiomatic formulation and he laid out a new approach to subjective probability through the concept of expected utility. Both Ramsey and Keynes were influenced by their Cambridge colleague William Ernest Johnson, whose early work on the Dirichlet priors for the multinomial is an antecedent to I.J. Good’s (77) approach for hierarchical Bayesian inference (see the discussion in Zabell (180)).

In an “objective” inverse probability vein, Ernest Lhoste (99) published a series of four articles comprising about 92 pages that appeared respectively in the May–August 1923 issues of the *Revue D’artillerie* in Paris. While some of this work was descriptive in nature, in the second article, arguing for a form of “indifference” in the spirit of Laplace, he developed vague prior distributions that represent little or no knowledge for the mean and variance of a normal distribution and for the probability of success for the binomial distribution. His results and reasoning were similar to those of Jeffreys (87) for the normal distribution and to those of Haldane for the binomial almost a decade later. Lhoste reasons that the prior density for the standard deviation, σ , should be the same as that for its reciprocal, and he sets

$$f(\sigma) \propto 1/\sigma \text{ for } \sigma > 0. \quad (3)$$

Thus he argues that lack of knowledge about σ should be the same as our lack of knowledge about $1/\sigma$. This is similar to Jeffreys’ (1961, page 119) invariance principle that states that prior information about σ should be the same as that for any power of σ . Broemeling and Broemeling (28) paraphrase Lhoste: “nothing distinguishes, a priori, $1/\sigma$ from σ ; if σ is indeterminate between two values, then $1/\sigma$ will be equally and in the same fashion indeterminate between the corresponding values of $1/\sigma$.” Thus what we now call “Jeffreys’ prior” appears to be another instance of Stigler’s Law. In the third article, Lhoste references Keynes (91) and makes clear that his contributions were not occurring in total isolation.

Beginning with his 1919 paper with Dorothy Wrinch (179), Harold Jeffreys adopted the “degree of reasonable belief” approach to probability, which was consistent with the approach in Keynes (91) but not with the “objective” ideas of Venn and others. They used this approach to assess the reasonableness of scientific theories, and in particular Jeffreys was focused on Eddington’s representation of relativity theory. Jeffreys’ 1931 book on *Scientific Inference* (86) represents a continuation of the collaboration with Wrinch, and in it he derived Gosset’s 1908 t -distribution result using inverse probability. Actually, he worked backward and showed that it corresponded to a prior proportional to $1/\sigma$. And in his description of the result he clearly missed the inverse probability underpinnings of Gosset’s work.

In the early 1930s, Jeffreys engaged in a published exchange with R.A. Fisher, beginning in the *Proceedings of the Cambridge Philosophical Society* and continuing in the *Proceedings of the Royal Society*. In this exchange they confronted one another on the meaning of probability, and on Fisher’s fiducial argument and Jeffreys’ inverse probability approach. Nothing seemed to be resolved, and Jeffreys resumed his critique

it is not widely recognized today as having substantial influence on the development of the subjective school.

of the Fisherian ideas in his 1939 book, *Theory of Probability* (87). That book provides Jeffreys' effort at an axiom system for probability theory. He then laid out the inverse probability approach of updating degrees of beliefs in propositions by use of probability theory—in particular Bayes'—theorem, to learn from experience and data. And he used an information-like invariance approach to derive “objective” priors that expressed ignorance or lack of knowledge in an effort to grasp the holy grail that had eluded statisticians since the days of Laplace. His examples remain among the most widely cited in the current “objective Bayesian” literature.

The Theory of Probability is a remarkable book, and it has been read and reread by many modern Bayesians. Jeffreys also taught a course in statistics at Cambridge. Dennis Lindley recalls (154): “Harold Jeffreys' lectures were attended by about six of us who had come back from the war and fancied ourselves as statisticians. That was the first time he had had to give the complete course of lectures. He was such a bad lecturer that previously all had given him up, but we stuck at them, and very rewarding they were.”

In Italy during the 1930s, in a series of papers in Italian, Bruno de Finetti gave a different justification for personal or subjective probability, introducing the notion of exchangeability and the implicit role of prior distributions (see especially (42)). But almost two decades were to pass before Savage was to build on these ideas and develop a non-frequentist alternative to the Kolmogorov axioms, and others were to exploit the concept of exchangeability to develop hierarchical Bayesian methods. The best known of these papers, de Finetti (42), was subsequently reprinted several times in English translation, e.g., in Kyburg and Smokler (94). A two-volume synthesis of his Bayesian ideas appeared in Italian in 1970 and in English translation in 1974 and 1975 (45; 46). These volumes and their description of de Finetti's ideas on finitely additive probabilities and non-conglomerability gave later statisticians the tools to study the implications of the use of improper prior distributions. Exchangeability was presaged by W.E. Johnson, who described the idea as “permutability” but did not develop it in the way that de Finetti did.

Meanwhile, in the United States, Bayes was not forgotten. W. Edwards Deming arranged for mimeo reprint of Bayes' essay, which was circulated by U.S. Department of Agriculture Graduate School during the late 1930s and early 1940s.

4.3 Statistical Developments in WWII

World War II marked a major turning point for statistics, both in Great Britain and in the United States.¹⁹ Several different groups of statisticians were assembled to deal with the war effort, and mathematicians were recruited to carry out statistical tasks. Wallis (176) and Barnard and Plackett (12) give excellent overviews of these activities.

¹⁹While there were statistical activities during World War II in other countries, none appear to have influenced the development of Bayesian methods in a fashion comparable to those that occurred in Great Britain and the United States. For a discussion of statistical activities in other countries see the overview in Fienberg (56).

Here I wish to focus on the simultaneous development of sequential analysis by Alan Turing, George Barnard, and Abraham Wald, as well as the personal links that were made that influenced some of the later course of Bayesian statistics.

I begin with Alan Turing, best known for his codebreaking work and his later contributions to computer science. I.J. Good (72), who was Turing’s statistical assistant at Bletchley Park during WWII, has described Turing’s statistical contributions during this period, which were primarily Bayesian. Good recalls that

[Turing] invented a Bayesian approach to sequential data analysis, using weights of evidence (though not under that name). A weight of evidence is the logarithm of a Bayes factor; for a Bayesian, this is the only possible definition, and the concept has been an obsession of mine ever since. . . . On one occasion I happened to meet George Barnard during the war, in London, and I confidentially [mentioned that we were using sequentially a Bayesian method in terms of what we now call “weights of evidence” (log-factors) for distinguishing between two hypotheses. Barnard said that, curiously enough, in his work for the Ministry of Supply, he was using something similar.] Thus Turing and Barnard invented sequential analysis independently of Wald. . . . Turing developed the sequential probability ratio test, except that he gave it a Bayesian interpretation in terms of the odds form of Bayes’ theorem. He wanted to be able to estimate the probability of a hypothesis, allowing for the prior probability, when information arrives piecemeal. When the odds form of Bayes’ theorem is used, it is unnecessary to mention the Neyman–Pearson lemma. One morning I asked Turing “Isn’t this really Bayes’ theorem?” and he said “I suppose so.” He hadn’t mentioned Bayes previously. Now, Harold Jeffreys with Dorothy Wrinch had previously published the odds form of Bayes’ theorem (without the odds terminology and without the sequential aspect), and Turing might have seen their work, but probably he thought of it independently.²⁰

As this quote implies, Turing was doing Bayesian work in depth, and Good was learning many of these ideas from him and later put them to work in his own research. It makes clear that George Barnard also developed the ideas of sequential analysis and after the war he published his results as well. In his 1946 paper (9) on the topic, he too noted the importance of Bayes’ theorem in the general problem of sampling inspection. Following the war, Good wrote his book *Probability and the Weighing of Evidence* (71) which was an effort to describe the inferential ideas that Good has used with Turing, and it was essentially completed in 1947 although not published for another three years. At the time, Good was unaware of de Finetti’s work.

Among the scientific efforts to support the U.S. war effort, W. Allen Wallis set up and directed a Statistical Research Group (SRG) at Columbia University during the Second World War. The staff included many mathematicians, statisticians, and economists who subsequently became leaders in their field including Milton Friedman, Frederick

²⁰ (8, pp. 10-11) as corrected by Good in personal correspondence.

Mosteller, L.J. Savage, Abraham Wald, and Jacob Wolfowitz (see Wallis (176)). SRG tackled a wide range of projects. For SRG, the idea of sequential sampling began with a question posed by a U.S. Navy captain about two firing procedures that were being tested to determine which procedure was “superior.” The captain indicated that, part way through the long test, he could often easily tell that one procedure was superior to the other. Wallis mulled this question over with Milton Friedman for a few days, and then they approached Abraham Wald with the problem. After a few months of work on the problem, Wald and his research group, which included Wolfowitz, developed and proved the theorem underlying the sequential probability ratio test, although unlike the versions of Turing and Barnard, his did not directly involve Bayes’ Theorem. This work was later published by Wald in 1947 in book form (174).

SRG did many other statistical tasks including work on quality control. Mosteller’s collaboration with Savage during this period was later to influence the development of Bayesian methods. And others were quick to follow up on the work of Wald. For example, David Blackwell credits a 1945 lecture by M.A. Girshick on sequential analysis that turned him into a statistician. Blackwell remembers Girshick recommending sampling “until you’ve seen enough.” In the lecture, Girshick announced a theorem that Blackwell thought was false. Blackwell devised a counter example and sent it to Girshick. It was wrong, but sending it was right. “Instead of dismissing it as the work of a crank,” Blackwell said, “he called and invited me to lunch” (excerpted from an ASA web profile (6)). Nine years and several papers later, in 1954, the two published their *Theory of Games and Statistical Decisions* (19). The book, while strongly influenced by the work of Wald and others, was still essentially frequentist, although others such as Morris DeGroot (48) later used ideas in it to develop the Bayesian notion of “sufficiency” for experiments.

Others built on the Wald approach as well. During the post war period, the Cowles Commission at the University of Chicago had two major research thrusts that were statistical in nature—estimation in simultaneous equations models and rational decision making (see Christ (31)). Those who worked on the latter topic included Kenneth Arrow, Herman Chernoff, M.A. (Abe) Girshick, and Herman Rubin, and later Roy Radner. Much of their work followed the Wald tradition, with Bayesian ideas being an integral component, but for frequentist purposes. Rubin (137) wrote about subjective probability ideas in several unpublished papers, as well as jointly with Arrow and Girshick on decision methods (5; 70), and he gave his own version of the von Neumann and Morgenstern axioms. Chernoff (30), leaning in part on the work of Savage for his foundations book, derived decision functions for minimizing maximum risk. He noted the seeming contradiction between his results and the subjective approach: “Theorem 3 . . . suggests that one may regard postulates 1-8 as an axiomatization of the ‘principle of insufficient reasoning.’ . . . Postulate 3 in particular is not compatible with a subjectivist approach.”

A similar decision-theory type of development was Robbins’ 1951 (133) paper on compound statistical decision problems, which introduced empirical Bayesian ideas, but from a strictly frequentist perspective. Turing had actually introduced empirical Bayes as a method as part of his wartime work, and Good developed these ideas further

in a 1953 paper (74), although it was not until the 1960s that these ideas entered the mainstream of Bayesian and frequentist thinking.

Quite separate from the statistical developments linked to World War II was the work of the physicists at Los Alamos, New Mexico, and elsewhere developing the atom bomb. But it was during this period that the ideas underlying the “Monte Carlo” method were developed, largely through the interactions of John von Neumann, Stanislaw Ulam, and Nicholas Metropolis (114). While we now think of one of the subsequent developments, the Metropolis algorithm, as a cornerstone for Bayesian computations, a careful reading of the initial paper (113) describing the method shows little or no evidence of Bayesian methodology or thinking (see also, Metropolis (112)).

Thus we see that, as statistics entered the 1950s, the dominant ideas had become frequentist, even though Bayes lurked in the background. As Dennis Lindley (106) has written: “When I began studying statistics in 1943 the term ‘Bayesian’ hardly existed; ‘Bayes’ yes, we had his theorem, but not the adjective.” And I.J. (Jack) Good (71), writing on the weighing of evidence using Bayes’ Theorem, in the third paragraph of the preface used the phrase “subjective probability judgments,” but nowhere in the book did he use the adjective “Bayesian.” He didn’t even use the phrase “Bayes factor,” although the concept appears under the label “factor,” following Turing’s suggestion to Good in a conversation in 1940. Good (73) also introduced precursors to hierarchical probability modeling as part of his post-WWII output.

5 The Neo-Bayesian Revival

5.1 First (?) Use of the Adjective “Bayesian”

According to David (39; 40; 41), the term “Bayesian” was first used in print by R.A. Fisher in the 1950 introduction to his 1930 paper on fiducial inference entitled “Inverse Probability,” as reprinted in his *Contributions to Mathematical Statistics* (65):

This short paper to the Cambridge Philosophical Society was intended to introduce the notion of “fiducial probability,” and the type of inference which may be expressed in this measure. It opens with a discussion of the difficulties which had arisen from attempts to extend Bayes’ theorem to problems in which the essential information on which Bayes’ theorem is based is in reality absent, and passes on to relate the new measure to the likelihood function, previously introduced by the author, and to distinguish it from the Bayesian probability *a posteriori*.

But in fact, as Edwards (51) notes, Fisher actually used the adjective earlier in *Contributions to Mathematical Statistics* (65), in his introduction to his 1921 paper “On the ‘probable error’ of a coefficient of correlation deduced from a small sample,” which was actually not reprinted in the volume:

In the final section this paper contains a discussion of the bearing of new ex-

act solutions of distributional problems on the nature of inductive inference. This is of interest for comparison with the paper on “Inverse probability,” published 1930, nine years later. In 1930 the notion of fiducial probability is first introduced using as an example the distribution found in this paper. In view of this later development the statement, “We can know nothing of the probability of hypotheses or hypothetical quantities,” is seen to be hasty and erroneous, in light of the different type of argument later developed. It will be understood, however, as referring only to the Bayesian probabilities *a posteriori*.

Fisher had spent decades defending his fiducial approach first from Jeffreys and others as distinct from inverse probability, and then from the adherents of the Neyman-Pearson school. Indeed his own arguments took on a more “Bayesian” flavor (although not subjective) as he fought against the Neyman-Pearson frequentists, especially in the last decade or so of his life. Edwards (51) gives a good over view of Fisher and Bayes’ Theorem and inverse probability.

In personal correspondence, Jack Good notes that “Bayesian” is now usually used to refer to a whole philosophy or methodology in which subjective or logical probabilities are used, and Fisher had a far more restricted notion in mind! But “Bayesian” is the word Fisher chose to use, and such negative usage of the term suggests that it might have been used similarly by others in previous oral exchanges. Similarly, Savage (138), in his review essay of Abraham Wald’s book on decision theory (Wald (175)), notes that “the problem of dealing with uncertainty when probability does not apply to the unknown states of the world is unBayesian, statistical theory.”

A search of JSTOR reveals no earlier usage in any of the main American and British statistical journals. The more comprehensive (in terms of journal coverage) permuted title index of Ross and Tukey (136) contains no use of the term in the title of articles until the 1960s. Similarly, an electronic search of *Current Index to Statistics* revealed only one “Bayesian” paper prior to the 1960s, which was also in the results of the JSTOR search.

5.2 Bayesian Inference Becomes “Bayesian”

Clearly Fisher’s derisory remark about Bayesian methods could not be responsible for the adoption of the term “Bayesian” and the widespread adoption of Bayesian methods in the second half of the twentieth century. Rather, at around the same time, there was a renewed interest in foundations and statistical decision theory, and this led to a melding of developments surrounding the role of subjective probability and new statistical tools for scientific inference and decisionmaking.

In the United States, the clear leader of this movement was Leonard Jimmie Savage. Savage had been part of the Columbia Statistical Group in WWII and later moved to the University of Chicago in 1949, where he began an intense reconsideration of the Kolmogorov approach to the axioms of probability. It was during this period that Savage

discovered the work of Bruno de Finetti. In his 1951 review of Wald’s book, Savage (138) points to the difficulty of the minimax approach, and he even cites de Finetti, but not in support of the subjective approach. Savage was grappling with ideas from von Neuman and Morgenstern (171) at the time, and he “discovered” and translated three early papers by Borel from the 1920s on game theory for publication in *Econometrica* (see (20; 22; 23)). He was also reading the work of Harold Jeffreys and Jack Good. Good (8) recalls,

The first time I met [Savage], he came to see me when I was in London. It was in 1951 or 1952 while he was working on his book. He’d been working in France and was visiting England briefly. He knew I’d written the 1950 book, so, perhaps on his way back to the U.S., he visited my home. Jimmie and I began corresponding after that. He pointed out an error in my very first paper on causality, when I sent him a draft. Later I saw him in Chicago. He was remarkably well read for a person with such bad eyesight.

It was Savage’s 1954 book (140) to which Good refers that set the stage for the neo-Bayesian revival. Thus it is surprising that not only is the term “Bayesian” absent from the book, but also that there is only one reference to Bayes, and that in connection with his theorem. In his first seven chapters, Savage laid out with mathematical rigor to rival that of Kolmogorov a series of axioms and theorems which could be derived from them, leading to the constructive methodology of maximizing expected utility, the ideas originally laid out in sketchier form by Ramsey (132) and von Neumann and Morgenstern (171) in their 1944 book, *Theory of Games and Economic Behavior*. His treatment relies heavily on ideas of personal probability from the work of Bruno de Finetti, in particular de Finetti (42; 43).²¹ Savage’s treatment of Jeffreys suggests that he did not regard his axiomatic approach as sufficiently mathematically rigorous to merit extended discussion and focus.²² In the second part of his book, Savage attempted to justify the frequentist ideas of Fisher and others using his axiomatic approach, but he later recognized the futility of such an effort. Looking back to this effort in 1970, Savage (144) described his thinking in 1954: “though interested in personal probability, . . . , not yet a personalistic Bayesian and . . . unaware of the likelihood principle.” At the same time, de Finetti (44) used the label in his paper, “La Notion de ‘Horizon Bayesien.’”²³

In the mid-1950s, there was what now appears to have been an amazing confluence of statisticians at the University of Chicago, in part drawn by the power and persuasiveness

²¹Savage and de Finetti were interacting and exchanging materials from the late 1940s onward, and there are clear cross-references in their work. For example, de Finetti (43) in his 1950 Berkeley symposium paper refers to an abstract of a paper presented by Savage in 1949 at the Econometric Society meeting and, in it, Savage refers to de Finetti’s ideas, referencing de Finetti (42).

²²Jeffreys had not used utility as part of his axiomatic structure, although he was aware of Ramsey’s work when he wrote his book.

²³In his review of this paper, I.J. Good (75) notes that it is “primarily an exposition of ideas previously put forward” in de Finetti (43). In that earlier paper, de Finetti does not use the adjective “Bayesian” but instead writes about “Bayes’ conclusions,” “Bayes’ premises,” “Bayes’ theory,” and the “Bayes’ position.” Thus the adjective had emerged in use as early as 1954 (the date when the paper was presented) and most certainly by 1956 when Good wrote his review.

of Savage. The faculty at the time included Harry Roberts and David Wallace; Morris (Morrie) DeGroot and Roy Radner were graduate students working with Savage as well. Among the visitors to the department during 1954-1955 were Dennis Lindley and Frederick Mosteller. Mosteller had worked with Savage in SRG at Columbia during WWII, and his 1948 paper on pooling data (115) was in the empirical Bayesian spirit, although it used frequentist criteria. He also did pioneering work with Nogee (116) on experimental measurement of utility. In 1952, Savage visited Cambridge, where Lindley was a member of the faculty. John Pratt came the next year:

I was at Chicago from September 1955 to June 1957. Jimmie's book had come out a year or two before, and I decided to read it. I sent him a note about misprints and such, on which he wrote replies. I have it, but it has no date. I got hung up on the de Finetti conjecture (p. 40) for a couple of weeks, but I read the rest of the book too. I had read and understood the result in Blackwell and Girshick,²⁴ but I think it stayed in another part of my brain, or maybe I didn't take it as meaning you should actually assess probabilities in principle. So perhaps it was Jimmie's book that made me realize how serious the whole matter is. (Pratt, personal communication)

My point here is that considerable prior interaction set the stage for the arrival of these individuals at the University of Chicago, and that Savage was the intellectual draw. It was during this period that Savage himself moved much closer to a fully subjective Bayesian position, and soon thereafter and continuing over the next decade, the adjective "Bayesian" began to appear in papers and books by all of these individuals, e.g., see Savage (143) and the classic expository article by Edwards, Lindman, and Savage (52). But Savage wrote about "Bayesian ideas" in a 1958 letter²⁵ to Dennis Lindley in connection with comments on a draft of Lindley (104).

A second and seemingly independent activity began a few years later at the Harvard Business School when Howard Raiffa joined forces with Robert Schlaifer to work on a Bayesian response to the frequentist theory of exponential family distributions, and they developed the notion of conjugate priors for this family. Raiffa was trained in the traditional mathematical statistics mold and his 1957 book with Luce (110), *Games and Decisions*, relied on the more traditional formulation of game theory.²⁶ Schlaifer was trained as a classical historian, and, according to Raiffa (see the interview with Fienberg (59)), when he was asked to teach statistics, "he read Fisher, Neyman and Pearson—not Wald and not Savage—and he concluded that standard statistical pedagogy did not address the main problem of a businessman: how to make decisions under uncertainty. Not knowing anything about the subjective/objective philosophical divide,

²⁴ "Blackwell and Girshick's 1954 book, *Theory of Games and Statistical Decisions*, a centrepiece of my graduate education, contains a complete and rigorous argument but presents it almost incidentally, in the midst of objectivistic exotica, and the clincher appears only in an exercise with no interpretation." Pratt (129) referring to Theorem 4.3.1 in Section 4.3 and Problem 4.3.1 from Blackwell and Girshick (19).

²⁵ Wallis (177, p. 21) quotes from this letter.

²⁶ Bayes appeared, but not yet in full Bayesian form. For example, on p. 312 they defined Bayes formula and on p. 313 they formulated the "Bayes decision rule against the a priori distribution."

he threw away the books and invented Bayesian decision theory from scratch.” In their collaboration, Raiffa and Schlaifer gave Bayesian definitions to frequentist notions such as “sufficiency,” which they said should come from the posterior distribution,²⁷ and they adapted Fisher’s and Jeffreys’ definitions of likelihood, noting the primary relevance of that part they called the “likelihood kernel.” How did they come to use the adjective “Bayesian” to describe their approach? Raiffa (59) has suggested that it was natural for Schlaifer to adopt the label “Bayesian” in a positive sense to describe his approach to probability reasoning and statistical methods. Schlaifer’s (147) 1959 introductory probability and statistics text not only uses the term but also advocates the use of Bayesian principles for business decisions.²⁸ By the time their classic book, *Applied Statistical Decision Theory*, appeared in 1961 (131), the label “Bayesian” showed up throughout, beginning with the preface. Shortly thereafter, John Pratt moved from Harvard’s Department of Statistics to the Harvard Business School and joined in this effort with a number of Raiffa and Schlaifer’s students and former students who had been trained as thorough Bayesians.

As I noted, shortly after these activities at Chicago and Harvard in the mid-1950s, “Bayesian” became the adjective of choice for both the proponents and the opponents of Bayesian methodology. As noted above, I.J. Good first used the term in a 1956 review of a paper by de Finetti (44) in *Mathematical Reviews*. But he began to use the label quite regularly after that, e.g., see Good (76), where he also introduced the phrase “Bayes factor,” and of course his important 1965 book (77). Lindley’s first “Bayesian” papers appeared in 1957 and 1958 (101; 102; 104), but like Savage his full acceptance of the Bayesian philosophy took many years to accomplish. Even in his two-volume 1965 text (105), Lindley focused on finding Bayesian solutions that resembled frequentist ones.

Savage likewise influenced many others. David Blackwell’s pioneering work with Girshick brought him close to Bayesian ideas, but “Jimmie convinced me that the Bayes approach is absolutely the right way to do statistical inference,” he has observed (6).

So how did Bayesian inference become “Bayesian”? Lindley (106) suggests,

Wald . . . had proved that the only decisions worth considering, technically the admissible solutions, were obtained by placing a probability distribution on the parameters about which a decision was to be made, and then using Bayes’ theorem. Moreover he called them Bayes solutions, using Bayes as an adjective, and although he did not use the term, it is but a short step to the proper adjectival form, Bayesian.

While the adjective first appears to have been used pejoratively, a small handful of statisticians embraced the notion of “Bayesian” to describe methods that revived inverse probability and imbued them with new methods and new mathematical foundations.

²⁷An earlier such definition of Bayesian sufficiency is due to Kolmogorov (92).

²⁸Savage (140), in the preface to the 1971 paperback edition of his book, refers to Schlaifer’s book and notes: “This is a welcome opportunity to say that his ideas were developed wholly independently of the present book, and indeed of other personalistic literature. They are in full harmony with the ideas in this book but are more down to earth and less spellbound by tradition.

Savage, Raiffa, and Schlaifer did not invent the eponym but, far more than Wald, their enthusiasm for and development of a new foundation for Bayesian methods encouraged others to adopt them and to use the Bayesian name.

In his 1958 paper, Good wrote the following:

By a neo-Bayesian or neo/Bayes-Laplace philosophy we mean one that makes use of inverse probabilities, with or without utilities, but without necessarily using Bayes' postulate of equiprobable or uniform initial distributions, and with explicit emphasis on the use of probability judgments in the form of inequalities. (76, p. 803)

Thus it seems apt to describe the 1950s as the era of the neo-Bayesian revival. But in many ways it took until 1962, and the publication of Allan Birnbaum's paper (18) on the likelihood principle as well as the "Savage volume" on *Foundations of Statistical Inference* (143), for the neo-Bayesian revival to become complete. The Savage volume was developed from a presentation Jimmie Savage gave in London in 1959 and included prepared discussions by Maurice Bartlett, George Barnard, David Cox, Egon Pearson, and C.A.B. Smith, and a more informal exchange that also included I.J. Good, Dennis Lindley, and others. Savage's introductory essay introduced many of the themes of modern Bayesian analysis including the role of the likelihood principle and the principle of "precise measurement" explored further in Edwards, Lindman and Savage (52). The Birnbaum paper was based on a special American Statistical Association discussion meeting held in late 1961, and its discussants included Savage, Barnard, Box, Good, Lindley, Pratt, and Dempster. In particular, Savage noted

I think that I, myself, came to ... Bayesian statistics ... seriously only through recognition of the likelihood principle; and it took me a year or two to make the transition. ... I can't know what everyone else will do, but I suspect that once the likelihood principle is widely recognized, people will not long stop at that halfway house but will go forward and accept the implications of personalistic probability for statistics. (18, p. 307)

Of course not all of the Birnbaum discussants were in agreement, either with him or with Savage. Irwin Bross, for example expressed his view that

the basic themes of this paper were well-known to Fisher, Neyman, Egon Pearson and others well back in the 1920's. But these men realized, as the author doesn't, that the concepts cannot be used for scientific *reporting*. So, they went on to develop confidence intervals in the 1930s ... The author here proposes to push the clock back 45 years, but at least this puts him ahead of the Bayesians, who would like to turn the clock back 150 years. (18, p. 310)

Lindley (107) recalls,

Savage [told] a good story about the [likelihood] principle. When he was first told it by George Barnard, he expressed surprise that anyone as brilliant as George could say something so ridiculous. Later he came to wonder how anyone could deny something so obviously correct.

The neo-Bayesian revival fused this renewed emphasis on the likelihood principle with Bayes’ Theorem and subjective probability as the mechanisms for achieving inferential coherence (c.f. Lindley (108)).

5.3 Departmental Homes for Statistics and the Neo-Bayesian Revival

Until the twentieth century, there were few university departments organized around separate disciplines, be they substantive or methodological, and when they began to emerge, there were few in statistics. The principal early exceptions were the enterprise led by Karl Pearson at University College London, founded shortly after the turn of the century, and the Statistics Laboratory at Iowa State College (now University), founded in the early 1930s. Nonetheless there were numerous chairs held by important figures in probability and statistics throughout this period, such as A.A. Chuprov in St. Petersburg, although these did not lead to the institutionalization of statistics outside of mathematics or economics.

Later in the U.S., Columbia University, George Washington University, the University of North Carolina, and North Carolina State University began separate departments. One can trace the growth of the identity of statistics as a separate identifiable discipline to the 1920s and 1930s and even 1940s, with the growing influence of Fisher and others. Nonetheless, even the creation of the Institute of Mathematical Statistics in the 1930s occurred at a time when most American statisticians resided in mathematics departments. Thus it should not be thought of as remarkable that most of those who contributed to the early developments of the evolution of Bayesian thinking were not necessarily identifiable as statisticians but rather as economists, mathematicians, scientists in the broadest of senses (e.g., Laplace and Gauss), philosophers, and physicists.

The institutionalization of statistics in the form of the creation of separate departments of statistics in the U.S. occurred after World War II, largely in the 1950s and 1960s, just at the time of the neo-Bayesian revival. For example, the department at Stanford was created in 1948, and a number of the new faculty had direct or indirect ties to the activities chronicled above. The department at the University of Chicago was created in 1949 (initially as a Committee), and was chaired by Jimmie Savage during the key years of the neo-Bayesian revival. Fred Mosteller, who was a visitor at Chicago in 1954-1955, returned to Harvard and was the first chair of the department there when it was created in 1957. George Box helped establish the department at the University of Wisconsin in 1960 and was soon joined by others of the Bayesian persuasion such as Irwin Guttman and George Tiao. Frank Anscombe, who worked on an alternative axiomatic treatment of subjective utility with Aumann (4), came to Yale in 1963 to help found its department and was joined by Jimmie Savage in 1964. Morrie DeGroot helped found the department at Carnegie Mellon in the mid 1960s. Not surprisingly,

all of these departments contributed to the neo-Bayesian revival and its aftermath; and without such homes, the revival may not have flourished and spawned the modern Bayesian school of statistics.

Nonetheless, the dominant philosophy in statistics departments in the United States remained frequentist for several decades, largely through the influence of Neyman and his University of California Berkeley colleagues and Fisher’s enduring impact on the statistics departments associated with applications in agriculture in U.S. land grant institutions.

5.4 Philosophical Linkages to Subjective Probability and the Bayesian Philosophy

There was a separate strain of intellectual activity on the foundations of probability that occurred more in philosophy than in statistics during the period from 1920 to 1960, only some of which I have mentioned. Since my story of the neo-Bayesian revival and the establishment of the “Bayesian” paradigm was largely built around those with close ties to mathematics and statistics, I did not give due credit to the developments in philosophy, some of which were influential to the Bayesian evolution. Rather than attempting to provide a complete chronicle of these developments, I will give a brief overview and mention some of the key figures.

It is not accidental that many of the key figures in the rise of subjective probability in the twentieth century were at Cambridge University and working in philosophy or on mathematical foundations linked to it.²⁹ While some of their interactions remain a matter of speculation, others are well documented. Richard Braithwaite, who posthumously edited Ramsey’s papers in the form of a book, was Ramsey’s contemporary at Cambridge and introduced him to Keynes in 1921, and they were all influenced, as I noted earlier, by W.E. Johnson. Jeffeys also interacted with Keynes and others at Cambridge. And Turing, Good, Barnard, and Lindley all fit within the legacy of this Cambridge tradition.

Despite Ramsey’s rejection of Keynes’ ideas on logical probabilities, they were later taken up by Rudolph Carnap, who made it the basis of the inductive logic published in 1950 in his *Logical Foundations of Probability* (29). Carnap, who was born in Germany and became part of the Vienna Circle of logical positivism before moving to the U.S. in the mid 1930s, was at the University of Chicago from 1936 to 1952. Savage (139) reviewed Carnap’s book, expressing dissatisfaction with his treatment and noting that de Finetti (whom Carnap did not cite) had a more suitable approach. Carnap nonetheless influenced many philosophers to adopt a Bayesian perspective, for example Richard Jeffrey. von Mises, who was trained in mathematics and physics in Vienna, was drawn into probability and statistics by his association with the Viennese Circle as well.

Ernest Nagel, who came to the U.S. at age 10 from Bohemia, wrote extensively on

²⁹It is also interesting but perhaps coincidental that two earlier statisticians in this story, Karl Pearson and R.A. Fisher, were students at Cambridge, and Fisher returned there in 1943 as professor of genetics.

the meaning of probability and, as noted above, was an advocate of the frequentist interpretation, (e.g., see his 1938 book (120)). Among his students at Columbia in the early 1950s were Henry Kyburg, Isaac Levi, and Patrick Suppes, all of whom wrote on the topic of subjective probability and had substantial contact with statisticians worrying about foundational issues. Suppes joined the faculty at Stanford University, where he interacted with Blackwell, Girshick, and Rubin, and he presented a paper on subjective probability and utility in decision-making at the 1955 Berkeley Symposium, published in 1956 (164). This paper presents an axiomatic treatment related to—but somewhat different from—Savage’s, and is notable in the present context because Suppes uses the adjective “Bayesian” once, in the discussion, thus supporting the notion that it was during this period that the term moved into usage by adherents to the subjective philosophy.³⁰ Suppes went on to make further contributions to the foundations of inference and measurement, e.g., in his collaboration with Luce, Krantz, and Tversky, and from 1960 onward had a cross-appointment in the Department of Statistics at Stanford.

Clearly there were direct and indirect contributions of many of these philosophers to the neo-Bayesian revival and its aftermath, but none except perhaps for Suppes seemed to be involved with the pivotal events that made Bayesian inference “Bayesian.”

5.5 The Stimulus of Fisher and Fiducial Inference

Fisher’s influence on the neo-Bayesian revival is not simply through his initial use of the adjective “Bayesian.” Following the publication of his 1956 book (66) on scientific inference,³¹ many statisticians attempted to make sense of his fiducial argument, especially his treatment of the Behrens-Fisher problem.³² This was deemed by many to be especially problematic because of his apparent need to integrate with respect to the distribution of incidental parameters, a problem tailor-made for Bayesian methods. Especially notable were the efforts of John Tukey, who devoted his 1958 IMS Wald Lectures to the topic.³³

Fisher’s 1956 book (66) also discussed the use of fiducial probability as *input to Bayes’ Theorem*. This leads to the type of inconsistency identified by Lindley (103), and which was in part was the basis for earlier correspondence with Fisher to which

³⁰Suppes recalls interactions with Herman Rubin during this period, especially in connection with the results in Savage and in Blackwell and Girshick, but not where the adjective “Bayesian” came from.

³¹As of July 5, 2005, the *Oxford English Dictionary Online* attributes the first usage of the adjective to Fisher’s 1956 book and the heading “Bayesian prediction”! Unpublished personal correspondence between I.J. Good and Dennis Lindley from 1994 suggests that they were the sources for this reference.

³²The Behrens-Fisher problem deals with making inferences about the difference between two normal means when the variances are unequal.

³³These lectures were not published, but Volume VI of Tukey’s *Collected Works* (166) includes the handout as well as a closely related unpublished paper from the previous year (167). In these Tukey covers much territory, with a special focus on the Fieller-Creasy and Behrens-Fisher problems and suggests the relevance of group invariance arguments later expounded by Fraser. The only reference to Bayesian alternatives comes indirectly with mention of Lindley (103), a paper which focuses on equivalence between fiducial and Bayesian solutions with uniform priors, with “Bayesian” being used throughout an an adjective.

Fisher replied: "...I have to thank you for some material you have written connected with my new book. I did not, of course expect you to approve of the latter, since it is evident that we speak somewhat different languages, and perhaps you will not expect me to understand what you mean by the 'relationship between fiducial probability and Bayes' theorem'."³⁴

Despite R.A. Fisher's death in 1962, fiducial inference remained alive through the 1960s, especially in the work of Don Fraser (67) and his students, although it slowly morphed into "structural inference." Jack Good (8) has reminded us that over the years many thought that Fisher's "fiducial argument was a failed attempt to arrive at a posterior degree of belief without mentioning a prior. (Harold Jeffreys pointed out what priors would patch up the argument.)" Similar remarks were made by others throughout the 1960s and 1970s, see e.g., Anscombe (3). Jimmie Savage (142) remarked that the Fisherian fiducial school's approach was "a bold attempt to make the Bayesian omelet without breaking the Bayesian eggs," a phrase that has been much repeated subsequently when others have used parts of the Bayesian approach without a full reliance on the specification of prior distributions and Bayesian updating. Seidenfeld (149) explores the link between Fisher's fiducial argument and Bayesian methods and points out that there are examples where the fiducial argument relies "on the sample space of possible observations to locate its Bayesian model," a property shared by many of the objective Bayesian methods for generating non-informative prior distributions. Barnard (11) in exploring some of these issues concludes that Fisher was not a Bayesian in the main current sense of the word.

6 Bayesian Inference Comes of Age in the 1960s

Bayesian statistics grew following the decade of the neo-Bayesian revival, both in terms of numbers of papers and numbers of authors. Here I chronicle a few of the major developments, but in doing so I omit reference to a large number of Bayesian authors who, like the present author, contributed to the literature.

Spurred on by their interactions when Mosteller visited Chicago in the mid-1950s, Fred Mosteller and David Wallace began a major collaboration to investigate who wrote the disputed Federalist papers. The Federalist papers were written mainly by Hamilton and Madison to help convince Americans to ratify the constitution in the 1780s. The authorship of most of the papers was clearly identifiable, but historians had been unable to establish whether a small number of them had been written by Hamilton or Madison. The work began in the summer of 1959 and, encouraged by Savage, they set to work on what was to become the first serious large-scale Bayesian empirical investigation involving heavy use of computers. Mosteller and Wallace produced posterior odds for each of the disputed papers, and in the process they reclaimed what we now know as Laplace's approximation for posterior densities. The work culminated in a 1963 paper (117), and their landmark 1964 book (118), which has since been reissued in an updated

³⁴Letter dated January 14, 1957, available at <http://www.library.adelaide.edu.au/digitised/fisher/corres/lindley/index.html>.

second edition.

Beginning around 1960 and running for about five years, Howard Raiffa and Robert Schlaifer ran a weekly seminar, the Decision Under Uncertainty seminar at the Harvard Business School. The seminar reported mostly on research in progress, and I remember attending these stimulating Monday afternoon gatherings as a first year graduate student in the Department of Statistics at Harvard in 1964. The next year, George Box and George Tiao spent their sabbatical leaves at the Harvard Business School, and Box lectured on Bayesian methods for time series from notes that became the Box-Jenkins ARIMA book (24). Box and Tiao had been working on their adaptation of Jeffreys’ method for choosing prior distributions and their Bayesian approach to robustness (25), which culminated in their 1973 Bayesian book (26). Another major project from the HBS seminar and the related research efforts was the Pratt, Raiffa, and Schlaiffer book on *Statistical Decision Theory* (130), whose 1965 “preliminary” version became a standard reference for those Bayesians lucky enough to have acquired a copy, even though it was not completed for another three decades.

Starting with the 1962 American Congressional election, John Tukey and others, working for the NBC television network, developed methodology to analyze early returns as they flowed in on election night. Projections were updated regularly as new data arrived. Decisions had to be made very quickly (to “beat” the other networks), but there was an extremely high premium on making the correct call on the winner. The statistical team had access to powerful computers in a laboratory owned by RCA, NBC’s parent company. David Brillinger (27) notes that Tukey described the work as “the best education in real-time statistics that anybody could have.” Data of several types were available: past history (at various levels, e.g., county), results of polls preceding the election, political scientists’ predictions, partial county returns flowing in during the evening, and complete results for selected precincts. The data of the analyses were, in many cases, “swings” (Republican percent minus Democratic percent) from sets of base values derived from past results and from political scientists’ opinions. Projecting turnout for locations with low partial returns turned out to be more difficult than projecting swings at lower levels of geography. The “improved” estimates that Tukey developed and programmed, with the collaboration of David Brillinger and David Wallace, were a form of empirical Bayes, where the past results were used to construct the prior distribution, much in the spirit of Bayesian analysis envisioned by Karl Pearson. Tukey often spoke of “borrowing strength” in this regard. The underlying model was in fact hierarchical Bayesian in structure and involved finite population structures, and the variance estimates, developed on a different basis by Brillinger and Wallace, were just as important as the point estimates. These tools were used with various refinements in elections from 1962 up through 1978. NBC stopped involving Tukey after the election of 1980 and relied instead on exit polls. Brillinger (27) notes that: “Tukey’s attitude to release of the techniques developed is worth commenting on. On various occasions members of his “team” were asked to give talks and write papers describing the work. When Tukey’s permission was sought, his remark was invariably that it was “too soon” and that the techniques were “proprietary” to RCA (the parent company for NBC).” In the 1970s, some essentially similar results appeared, but were published

by Alastair Scott and Fred Smith (148) and not members of the NBC team.³⁵ Dennis Lindley and Adrian Smith (109) presented related results for hierarchical modeling in the normal linear model. Thus there is no published record of this earlier large-scale Bayesian hierarchical modeling work.

As I noted earlier, Robbins published his first paper on empirical Bayesian methods in 1951. He continued to develop these ideas in two later papers published in 1956 and 1961 (134; 135), but this work remained frequentist in nature. Jack Good (74) and Toulmin (81) implemented a Bayesian version of empirical Bayes for a specific problem, following up on ideas suggested during World War II by Turing, e.g., see Good's introduction to a reprinting of Robbins' 1956 paper (80). During the same period of time, Charles Stein demonstrated the inadmissibility of the usual estimate of the multivariate normal mean and showed how to get improved estimates from approximate Bayesian arguments where sample quantities were used to estimate prior distribution parameters. Efron and Morris (54) used the name empirical Bayes to describe variations on the methodology of Stein, and over the years their approach has come close to integration with fully Bayesian methods because of links to hierarchical models, especially in the case of exponential family distributions. But recent work using the voluminous data associated with microarrays in genetics has revived interest in Robbins' earlier instantiation of empirical Bayes with nonparametric estimation of the prior distribution.

In 1960, Dennis Lindley moved to a professorship at University College in Aberystwyth, Wales, where he built a Bayesian group. Several years later he moved again to University College London, where, in very short order, he established a strong Bayesian department that was to radically change the face of Bayesian work in the United Kingdom and far beyond. Jimmie Savage moved, first to the University of Michigan and then to Yale University, and continued to exert strong influence on the thinking of both colleagues and students (e.g., see the comments in Kadane (88)). His remarkable 1970 Fisher lecture, published posthumously six years later through the editorial efforts of John Pratt (145), kept a spellbound audience focused on Fisher's contributions to statistics and Savage's view of them for well beyond the allotted time for the session. And Jack Good moved from England to the United States, but worked largely in isolation from other Bayesians.

Shortly after Savage came to Michigan, Ward Edwards, who had done work on subjective expected utility, and one of his graduate students, Harold Lindman, developed the first accessible nontechnical exposition about the robustness of Bayesian specifications relative to the prior, in their citation classic 1963 paper (52). Around this time Edwards began an annual series of Bayesian research conferences primarily for psychologists which have continued to the present day (the 42nd one was held in January, 2004). Savage attended some of the early meetings.

³⁵ Alastair Scott was David Wallace's Ph.D. student at the University of Chicago from 1962 to 1965 but according to Scott they never discussed these ideas, I presume in keeping with Tukey's notion that they were proprietary! David Brillinger and Scott were colleagues at the London School of Economics while these ideas were gestating and was aware of the Scott-Smith work. But he too was constrained from saying anything because of Tukey.

At the National Institutes of Health (NIH), Jerry Cornfield became interested in Bayesian ideas during the late 1950s. He was a discussant of Birnbaum (18) and was, at the time, highly supportive of the likelihood approach. Around this time he was engaged in foundational discussions with Jimmie Savage, and he visited Michigan in the early 1960s to give a talk on his own developing ideas about Bayesian inference, partly to help formulate his thinking, but also to meet with Savage. Seymour Geisser joined Cornfield at NIH and often had luncheon conversations about Bayesian ideas (see Christensen and Johnson (32)). These led to their joint 1962 paper (68) on the multivariate normal distribution, where they contrasted Bayesian methods with improper priors to fiducial solutions attributable to Fisher. Later Geisser was one of the discussants (along with Hartley, Kempthorne, and Herman Rubin) of Cornfield’s 1969 paper (34) in which he led the reader step by step along the path that compelled him to be a Bayesian (condition, recognizable subsets, betting, and the meaning of probability) along with a report on a serious epidemiological application. Both continued to stress Bayesian ideas through the remainders of their careers, with Geisser stressing the importance of observables (in the tradition of de Finetti) and predictive inference.

During the 1950s, a casualty actuary, Arthur Bailey (7), wrote about the link between what was then known as credibility theory and “the Laplace generalization of Bayes’ rule”—i.e., credibility theory added informative prior distributions. In 1964, Mayerson (111) made the connection with Raiffa and Schlaiffer’s work on conjugate prior distributions. This then led to subsequent Bayesian work on credibility theory, most notably by H. Bühlmann and by W.S. Jewell who wrote on the link to exponential family theory. As an interesting side note, de Finetti, who was professor of actuarial science in Rome, participated in a number of international conferences during the 1950s and 1960s but seemed not to be involved directly in any of the discussions regarding credibility theory and Bayesian statistics.

Arnold Zellner recalls participating in the early 1960s in a University of Wisconsin seminar devoted to the reading of Jeffreys’ book, along with George Box, Mervyn Stone (as a visitor), Irwin Guttman, George Tiao, and others. He then collaborated with George Tiao (165), and when he moved to the University of Chicago, he continued to advocate the use of Bayesian ideas in econometrics and extended the work of Jeffreys on the generation of “non-informative” invariant priors, especially in his 1971 book (183). It was during this period that Zellner first organized the semi-annual NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics at the University of Chicago. Before long, statistics groups at other universities also hosted the seminar, and for almost two decades it was a major locus for the presentation of Bayesian research.

7 The Modern Bayesian Era

In this paper, I have attempted to give an overview of key Bayesian developments, beginning in England with Bayes’ posthumously published 1763 paper, emerging in France soon thereafter with Laplace, and spreading to other parts of Europe in the nineteenth century. But the rise of subjective probability in the twentieth century

and the statistical efforts related to it occurred largely in England and then later in the U.S., with notable exceptions such as the contributions of de Finetti. Activities surrounding World War II had a profound influence on probability and statistics, not only in terms of the emigration of a number of key figures from Europe to the U.S. (e.g., Carnap, Neyman, von Mises, von Neumann, Wald) but also in terms of the impact that the application of statistics had on the development of mathematical statistics and foundations. The work during the 1950s, a period which I have called the neo-Bayesian revival when “Bayesian” emerged as the label of choice for those who advocated Bayesian methods, was dominated by statisticians in the U.S. and England, as were the developments of the 1960s. The blossoming of the Bayesian approach around the world in the three decades that followed is rooted in this history.

In some ways, this paper fills in a number of Bayesian details to my earlier overview of the history of statistics (57), and it also brings that history forward by a couple of decades, at least in part. For an overview of some recent research in Bayesian methods and foundations see the 2004 special issue of *Statistical Science*.

A major focal point in the development of Bayesian statistics since the early 1970s has been the quadrennial Valencia Symposium on Bayesian Statistics organized by Jose Bernardo, the proceedings of which have set a high standard for Bayesian publications. The introduction of Monte Carlo Markov chain methods to the Bayesian world in the late 1980s made possible computations that others had only dreamed of two decades earlier, when the work on the Federalist papers or election predictions was viewed as a Herculean feat. Bayesian methods have spread rapidly into areas of application and have been championed by practicing scientists, e.g., in computer science and machine learning. In 1992, both the International Society for Bayesian Analysis and the Bayesian Statistical Science Section of the American Statistical Association were founded (and since then have held many successful meetings), and ISBA has just launched the online statistical journal in which this article appears. And the quest for the holy grail of a method for deriving “objective” non-informative priors to express ignorance continues to the present day, aided in part by a continuing series of workshops on the topic.

Over the period since the neo-Bayesian revival, many have looked for syntheses of Bayesian and non-Bayesian ideas and methods, e.g., see Lindley (105), Pratt (128) and Good (78) for early examples, as well as the many recent contributions by Jim Berger and his collaborators. Moreover, the unity of statistics has become a theme of many papers, beginning as early as George Barnard’s 1971 presidential address to the Royal Statistical Society (10) and continuing to the present day with Bradley Efron’s 2004 presidential address to the American Statistical Association (53).

These developments clearly are not all attributable to the use of the adjective “Bayesian.” But they have meant that more and more statisticians and users of statistics have an interest in the origin of the label that describes the ideas and methods upon which they rely.

8 Some Sources

Many of the papers cited here are available from online electronic archives such as those maintained by JSTOR, and from online web resources of a handful of technical publishers. The collections edited by Kyburg and Smokler are especially valuable sources. The first edition of their book includes an excerpt from Venn, a translation of Borel (21), Ramsey (132), a translation of de Finetti (42), a paper by Koopman, and Savage (142). The second edition includes an overlapping and updated set of papers including later ones by de Finetti, Good and Savage. Several of the classic books from the crucial period of the 1950s referenced here are still in print in special paperback editions, and several papers cited here have been reprinted in Good (79) and Savage (146). The collection edited by Heyde and Seneta (84) contains excellent succinct biographies of many of the historical figures mentioned here, but none of the key figures in the neo-Bayesian revival, as the collection covers only those born prior to the twentieth century.

Bibliography

- [1] Aldrich, J. (1997). “R.A. Fisher and the Making of Maximum Likelihood 1912–1922.” *Statistical Science*, 12: 162–176. 7
- [2] — (2002). “How Likelihood and Identification Went Bayesian.” *International Statistical Review*, 70. 2
- [3] Anscombe, F. J. (1961). “Bayesian Statistics.” *American Statistician*, 15: 21–24. 23
- [4] Anscombe, F. J. and Aumann, R. J. (1963). “A Definition of Subjective Probability.” *Annals of Mathematical Statistics*, 34: 199–205. 20
- [5] Arrow, K. J., Girshick, M. A., , and Rubin, H. (1949). “Bayes and Minimax Solutions of Sequential Decision Problems.” *Econometrica*, 17: 213–244. 13
- [6] Association, A. S. (2003). “David Blackwell. Go Bayes!” *Amstat News*, 6–7. 13, 18
- [7] Bailey, A. L. (1950). “Credibility Procedures, Laplace’s Generalization of Bayes’ Rule and the Combination of Collateral Knowledge with Observed Data.” In *Proceedings of the Casualty Actuarial Society*, volume 37, 7–23. 26
- [8] Banks, D. L. (1996). “A Conversation with I.J. Good.” *Statistical Science*, 11: 1–19. 12, 16, 23
- [9] Barnard, G. A. (1946). “Sequential Tests in Industrial Statistics.” *Supplement to the Journal of the Royal Statistical Society*, 8: 1–26. 12
- [10] — (1972). “The Unity of Statistics.” *Journal of the Royal Statistical Society, Series A*, 135: 1–15. 27

- [11] — (1987). “R. A. Fisher—A True Bayesian?” *International Statistical Review*, 55: 183–189. [23](#)
- [12] Barnard, G. A. and Plackett, R. L. (1985). “Statistics in the United Kingdom, 1939–45.” In Atkinson, A. C. and Fienberg, S. E. (eds.), *A Celebration of Statistics: The ISI Centenary Volume*, 31–56. New York: Springer-Verlag. [11](#)
- [13] Barnett, V. (1973). *Comparative Statistical Inference*. Wiley. [9](#)
- [14] Bayes, T. (1763). “An Essay Towards Solving a Problem in the Doctrine of Chances.” *Philosophical Transactions of the Royal Society of London*, 53: 370–418. [3](#)
- [15] Bellhouse, D. R. (2002). “On Some Recently Discovered Manuscripts of Thomas Bayes.” *Historia Mathematica*, 29: 383–394. [4](#)
- [16] — (2004). “The Reverend Thomas Bayes FRS: A Biography to Celebrate the Tercentenary of his Birth (with discussion).” *Statistical Science*, 3–43. [2](#), [3](#)
- [17] Bienaymé, I. (1838). “Mémoire sur la Probabilité des Résultats Moyens des Observations; Démonstration Directe de la Règle de Laplace.” *Mém. Pres. Acad. Roy. Sci. Inst. France*, 5: 513–558. [5](#)
- [18] Birnbaum, A. (1962). “On the Foundations of Statistical Inference (with discussion).” *Journal of the American Statistical Association*, 57: 269–326. [19](#), [26](#)
- [19] Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley. [13](#), [17](#)
- [20] Borel, E. (1921). “La Théorie du Jeu et les Équations Intégrales à Noyau Symétrique.” *Comptes Rendus de l’Académie des Sciences*, 173: 1304–1308. Translated by L. J. Savage as “The Theory of Play and Integral Equations with Skew Symmetric Kernels.” *Econometrica*, **21**, 97–100. [16](#)
- [21] — (1924). “À Propos d’un Traité de Probabilités.” *Revue Philosophique*, 98: 321–336. Translated as “A Propos of a Treatise on Probability.” In Kyburg, H. E. and Smokler, H. E., eds. (1964). *Studies in Subjective Probability*. Wiley, New York, 45–60. [9](#), [28](#)
- [22] — (1924). “Sur les Jeux où Interviennent L’hasard et L’habileté des Joueurs.” In *Théorie des Probabilités*, 204–224. Librairie Scientifique, J. Herman. [16](#)
- [23] — (1927). “Sur les Systèmes de Formes Linéaires à Déterminant Symétrique Gauche et la Théorie Générale du Jeu. from Algèbre et Calcul des Probabilités.” *Comptes Rendus de l’Académie des Sciences*, 173: 52–53. [16](#)
- [24] Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day. [24](#)
- [25] Box, G. E. P. and Tiao, G. C. (1961). “A Further Look at Robustness Via Bayes’ Theorem.” *Biometrika*, 49: 419–432. [24](#)

- [26] — (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley. 24
- [27] Brillinger, D. A. (2002). “John W. Tukey: His Life and Professional Contributions.” *Annals of Statistics*, 30: 1535–1575. 24
- [28] Broemeling, L. and Broemeling, A. (2003). “Studies in the History of Probability and Statistics: XLVIII. The Bayesian Contributions of Ernest Lhoste.” *Biometrika*, 90: 728–731. 10
- [29] Carnap, R. (1950). *Logical Foundations of Probability*. University of Chicago Press. 21
- [30] Chernoff, H. (1954). “Rational Selection of Decision Functions.” *Econometrica*, 22: 422–443. 13
- [31] Christ, C. F. (1994). “The Cowles Commission’s Contributions to Econometrics at Chicago, 1939-1955.” *Journal of Economic Literature*, 32: 30–59. 13
- [32] Christensen, R. and Johnson, W. (2004). “Conversation with Seymour Geisser.” *Statistical Science*. 26
- [33] Condorcet, M.-J.-A.-N. d. C. (1785). *Essai sur L’application de L’analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*. Imprimerie Royale. 5
- [34] Cornfield, J. (1969). “The Bayesian Outlook and Its Application (with discussion).” *Biometrics*, 25: 617–657. 26
- [35] Cournot, A. A. (1984). *Exposition de la Theorie des Chances et des Probabilités*. Librairie J. Vrin. 6
- [36] Dale, A. I. (1999). *A History of Inverse Probability from Thomas Bayes to Karl Pearson*. Springer-Verlag. 2, 5, 37
- [37] — (2003). *Most Honourable Remembrance: The Life and Work of Thomas Bayes*. Springer-Verlag. 3, 4
- [38] Daston, L. (1994). “How Probabilities Came to be Objective and Subjective.” *Historia Mathematica*, 21: 330–344. 6
- [39] David, H. A. (1995). “First (?) Occurrence of Common Terms in Mathematical Statistics.” *American Statistician*, 49: 121–133. 14
- [40] — (1998). “First (?) Occurrence of Common Terms in Probability and Statistics – A Second List, with Corrections.” *American Statistician*, 52: 36–40. 14
- [41] — (2001). “Appendix B.” In David, H. A. and Edwards, A. W. F. (eds.), *Annotated Readings in the History of Statistics*. Springer-Verlag. 3, 14

- [42] de Finetti, B. (1937). “La Prévision: Ses Lois Logiques, Ses Sources Subjectives.” *Annales de l’Institut Henri Poincaré*, 7: 1–68. Translated as “Foresight: Its Logical Laws, Its Subjective Sources,” in Kyburg, H. E. and Smokler, H. E. eds., (1964). *Studies in Subjective Probability*. Wiley, New York, 91–158. (Corrections appear in (de Finetti, 1972, p. xvii)). [11](#), [16](#), [28](#)
- [43] — (1951). “Recent Suggestions for the Reconciliation of Theories of Probability.” In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, 217–225. Berkeley: University of California Press. [16](#)
- [44] — (1955). “La Notion de ‘Horizon Bayésien.’” In Thone, G. (ed.), *Colloque Sur l’Analyse Statistique, Bruxelles*, 57–71. Paris: Liège and Masson & Cie. [16](#), [18](#)
- [45] — (1974). *Theory of Probability*, volume I. Wiley. [11](#)
- [46] — (1975). *Theory of Probability*, volume II. Wiley. [11](#)
- [47] De Morgan, A. (1837). “Review of Laplace’s *Théorie Analytique des Probabilités*. (3rd Edition).” *Dublin Review*, 2, 3: 338–354, 237–248. [5](#)
- [48] DeGroot, M. A. (1970). *Optimal Statistical Decisions*. McGraw-Hill. [13](#)
- [49] Edgeworth, F. Y. (1883). “The Method of Least Squares.” *Philosophical Magazine*, 5th series, 34: 190–204. [6](#)
- [50] Edwards, A. W. F. (1997). “What Did Fisher Mean by ‘Inverse Probability’ in 1912–1922?” *Statistical Science*, 12: 177–184. [7](#)
- [51] — (2004). “Comment on Bellhouse, David R. ‘The Reverend Thomas Bayes FRS: A Biography to Celebrate the Tercentenary of his Birth’.” *Statistical Science*, 19: 34–37. [14](#), [15](#)
- [52] Edwards, W., Lindman, H., and Savage, L. J. (1963). “Bayesian Statistical Inference for Psychological Research.” *Psychological Review*, 70: 193–242. [17](#), [19](#), [25](#)
- [53] Efron, B. (2005). “Bayesians, Frequentists, and Scientists.” *Journal of the American Statistical Association*, 100. [27](#)
- [54] Efron, B. and Morris, C. (1973). “Stein’s Estimation Rule and its Competitors—An Empirical Bayes Approach.” *Journal of the American Statistical Association*, 68: 117–130. [25](#)
- [55] Fechner, G. T. (1897). *Kollektivmasslehre*. Engelmann. [6](#)
- [56] Fienberg, S. E. (1985). “Statistics Developments in World War II: An International Perspective.” In Atkinson, A. C. and Fienberg, S. E. (eds.), *A Celebration of Statistics: The ISI Centenary Volume*, 25–30. Springer-Verlag. [11](#)
- [57] — (1992). “A Brief History of Statistics in Three and One-half Chapters: A Review Essay.” *Statistical Science*, 7: 208–225. [27](#)

- [58] — (1999). “Statistics in the Social World,’ a review of *The Politics of Large Numbers: A History of Statistical Reasoning*, by Alain Desrosières.” *Science*, 283: 2025. [2](#)
- [59] — (2005). “Interview with Howard Raiffa.” *Statistical Science*. [17](#), [18](#)
- [60] Fienberg, S. E. and Lazar, N. (2001). “William Sealy Gosset: 1876-1937.” In Heyde, C. and Seneta, E. (eds.), *Statisticians of the Centuries*. New York: Springer-Verlag. [7](#)
- [61] Fienberg, S. E. and Tanur, J. M. (1996). “Reconsidering the Fundamental Contributions of Fisher and Neyman on Experimentation and Sampling.” *International Statistical Review*, 64: 237–253. [8](#)
- [62] Fisher, R. A. (1922). “On the Mathematical Foundations of Theoretical Statistics.” *Philosophical Transactions of the Royal Society of London, Series A*, 222: 309–368. [4](#), [8](#)
- [63] — (1925). *Statistical Methods for Research Workers*. Oliver and Boyd. [8](#)
- [64] — (1930). “Inverse Probability.” *Proceedings of the Cambridge Philosophical Society*, 26: 528–535. [8](#)
- [65] — (1950). *Contributions to Mathematical Statistics*. Wiley. [14](#)
- [66] — (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd. [22](#)
- [67] Fraser, D. A. S. (1968). *The Structure of Inference*. Wiley. [23](#)
- [68] Geisser, S. and Cornfield, J. (1963). “Posterior Distributions for Multivariate Normal Parameters.” *Journal of the Royal Statistical Society, Series B*, 25: 368–376. [26](#)
- [69] Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., , and Krüger, L. (1989). *The Empire of Chance: How Probability Changed Everyday Life*. Cambridge University Press. [8](#)
- [70] Girshick, M. A. and Rubin, H. (1952). “A Bayes Approach to a Quality Control Model.” *Annals of Mathematical Statistics*, 23: 114–125. [13](#)
- [71] Good, I. (1950). *Probability and the Weighing of Evidence*. Charles Griffin. [12](#), [14](#)
- [72] — (1979). “Studies in the History of Probability and Statistics: XXXVII. Turing’s Statistical Work in World War II.” *Biometrika*, 66: 393–396. [12](#)
- [73] Good, I. J. (1952). “Rational Decisions.” *Journal of the Royal Statistical Society, Series B*, 14: 107–114. [14](#)
- [74] — (1953). “The Population Frequencies of Species and the Estimation of Population Parameters.” *Biometrika*, 40: 237–264. [14](#), [25](#)

- [75] — (1956). “Review of B. de Finetti (1954).” *Mathematical Reviews*, 17: 633. 16
- [76] — (1958). “Significance Tests in Parallel and in Series.” *Journal of the American Statistical Association*, 53: 799–813. 18, 19
- [77] — (1965). *The Estimation of Probabilities*. The M.I.T. Press. 10, 18
- [78] — (1976). “The Bayesian Influence, or How to Sweep Subjectivism Under the Carpet.” In Hooker, C. A. and Harper, W. (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 2, 125–174. Dordrecht, Holland: D. Reidel. 27
- [79] — (1983). *Good Thinking: The Foundations of Probability and its Applications*. University of Minnesota Press. 28
- [80] — (1992). “Introduction to Robbins (1955), An Empirical Bayes Approach to Statistics.” In Johnson, N. L. and Kotz, S. (eds.), *Breakthroughs in Statistics, Vol. I: Foundations and Basic Theory*, 379–387. Springer-Verlag. This is in fact an introduction to Robbin’s 1956 paper (134). 25
- [81] Good, I. J. and Toulmin, G. (1956). “The Number of New Species, and the Increase in Population Coverage, When a Sample is Increased.” *Biometrika*, 43: 45–63. 25
- [82] Hald, A. (1998). *A History of Mathematical Statistics From 1750 to 1930*. Wiley. 2, 4
- [83] Heyde, C. C. and E., S. (1977). *I. J. Bienaymé: Statistical Theory Anticipated*. Springer-Verlag. 5, 6
- [84] — (2001). *Statisticians of the Centuries*. Springer-Verlag. 28
- [85] Howie, D. (2002). *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*. Cambridge University Press. 5
- [86] Jeffreys, H. (1931). *Scientific Inference*. Cambridge University Press. 10
- [87] — (1939). *Theory of Probability*. Oxford University Press. 7, 10, 11
- [88] Kadane, J. B. (2001). “Jimmie Savage: An Introduction.” *ISBA Bulletin*, 8: 5–6. 25
- [89] Kapteyn, J. C. (1903). *Skew Frequency Curves in Biology and Statistics*. Noordhoff. 4
- [90] Kasner, E. and Newman, J. R. (1940). *Mathematics and the Imagination*. Simon and Schuster. 1
- [91] Keynes, J. M. (1921). *A Treatise on Probability*, volume 8. St Martin’s. 9, 10
- [92] Kolmogorov, A. (1933). “Sur L’estimation Statistique des Parametres de la Loi de Gauss.” *Bulletin of the Academy of Science URSS Ser. Math.*, 6: 3–32. 18

- [93] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer-Verlag. 2, 8
- [94] Kyburg, H. E. and Smokler, H. E. (eds.) (1964). *Studies in Subjective Probability*. New York: Wiley. Second revised edition, Krieger, Garden City, 1980. 11
- [95] Lacroix, S. F. (1816). “Traité Élémentaire du Calcul des Probabilités.” 5
- [96] Laplace, P.-S. (1774). “Mémoire sur la Probabilité des Causes par les Événements.” *Mémoires de Mathématique et de Physique Présentés à l’Académie Royale des Sciences, Par Divers Savans, & Lûs dans ses Assemblées*, 6: 621–656. 3
- [97] — (1812). *Théorie Analytique des Probabilités*. Courcier. 5
- [98] Lévy, P. (1925). *Calcul de Probabilités*. Gauthier-Villars. 9
- [99] Lhoste, E. (1923). “Le Calcul des Probabilités Appliqué a L’artillerie.” *Revue D’artillerie*, 91, 92: 405–423 and 516–532, 58–82 and 152–179. 10
- [100] Lindley, D. V. (1953). “Statistical Inference.” *Journal of the Royal Statistical Society, Series B*, 16: 30–76. 9
- [101] — (1957). “Fiducial Distributions and Bayes’ Theorem.” *Journal of the Royal Statistical Society, Series B*, 20: 102–107. 18
- [102] — (1957). “A Statistical Paradox.” *Biometrika*, 44: 187–192. 18
- [103] — (1958). “Fiducial Distributions and Bayes’ Theorem.” *Journal of the Royal Statistical Society, Series B*, 20: 102–107. 22
- [104] — (1958). “Professor Hogben’s ‘Crisis’—A Survey of the Foundations of Statistics.” *Applied Statistics*, 7: 186–198. 17, 18
- [105] — (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 1: Probability. Part 2: Inference*. Cambridge University Press. 18, 27
- [106] — (2000). “What is a Bayesian?” *ISBA Bulletin*, 1: 7–9. 2, 14, 18
- [107] — (2004). “Bayesian Thoughts (An Interview with Helen Joyce).” *Significance*, 1: 73–75. 19
- [108] — (2004). “That Wretched Prior.” *Significance*, 1: 85–87. 20
- [109] Lindley, D. V. and Smith, A. F. M. (1972). “Bayes Estimates for the Linear Model (with discussion).” *Journal of the Royal Statistical Society, Series B*, 34: 1–44. 25
- [110] Luce, D. and Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*. Wiley. 17

- [111] Mayerson, A. L. (1964). “A Bayesian View of Credibility.” *Proceedings of the Casualty Actuarial Society*, 51: 85–104. 26
- [112] Metropolis, N. (1987). “The Beginning of the Monte Carlo Method.” *Los Alamos Science*, 125–130. 14
- [113] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). “Equations of State Calculations by Fast Computing Machines.” *Journal of Chemical Physics*, 21: 1087–1092. 14
- [114] Metropolis, N. and Ulam, S. (1949). “The Monte Carlo Method.” *Journal of the American Statistical Association*, 44: 335–341. 14
- [115] Mosteller, F. (1948). “On Pooling Data.” *Journal of the American Statistical Association*, 43: 231–242. 17
- [116] Mosteller, F. and Nogee, P. (1951). “An Experimental Measurement of Utility.” *Journal of Political Economy*, 59: 371–404. 17
- [117] Mosteller, F. and Wallace, D. L. (1963). “Inference in an Authorship Problem.” *Journal of the American Statistical Association*, 58: 275–309. 23
- [118] — (1964). *Inference and Disputed Authorship: The Federalist*. Addison-Wesley. 23
- [119] Nagel, E. (1936). “The Meaning of Probability (with discussion).” *Journal of the American Statistical Association*, 31: 10–30. 8, 9
- [120] — (1938). *The Principles of the Theory of Probability*. University of Chicago Press. 22
- [121] Neyman, J. (1937). “Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability.” *Philosophical Transactions of the Royal Society, Series A*, 236: 333–380. 9
- [122] of Mathematical Statistics, I. (1988). “The Reverend Thomas Bayes, F.R.S.—1701?-1761.” *IMS Bulletin*, 17: 276–278. 4
- [123] Pearson, K. (1892). *The Grammar of Science*. Walter Scott. 7
- [124] — (1907). “On the Influence of Past Experience on Future Expectation.” *Philosophical Magazine*, 13: 365–378. 7
- [125] Pfanzagl, J. and Sheynin, O. (1996). “Studies in the History of Probability and Statistics: XLIV. A Forerunner of the t -distribution.” *Biometrika*, 83: 891–898. 7
- [126] Porter, T. (1986). *The Rise in Statistical Thinking: 1820–1900*. Princeton University Press. 2

- [127] — (2004). *Karl Pearson: The Scientific Life in a Statistical Age*. Princeton University Press. [1](#)
- [128] Pratt, J. W. (1965). “Bayesian Interpretation of Standard Inference Statements.” *Journal of the Royal Statistical Society. Series B*, 27. [27](#)
- [129] — (1995). “Foreword.” In Eeckhoudt, L. and Gollier, C. (eds.), *Risk: Evaluation, Management and Sharing*. New York: Harvester Wheatsheaf. [17](#)
- [130] Pratt, J. W., Raiffa, H., and Schlaifer, R. (1965). *Introduction to Statistical Decision Theory—Preliminary Edition*. McGraw-Hill. [24](#)
- [131] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Division of Research Graduate School of Business Administration, Harvard University. [18](#)
- [132] Ramsey, F. P. (1926). “Truth and Probability.” Published in 1931 as *Foundations of Mathematics and Other Logical Essays*, Ch. VII, pp. 156–198. Edited by R.B. Braithwaite. Kegan, Paul, Trench, Trubner & Co., London. Reprinted in Kyburg, H. E. and Smokler, H. E. eds. (1964). *Studies in Subjective Probability*. Wiley, New York, 61-92. [9](#), [16](#), [28](#)
- [133] Robbins, H. (1951). “Asymptotically Subminimax Solutions of Compound Statistical Decision Problems.” In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, 131–148. Berkeley. [13](#)
- [134] — (1956). “An Empirical Bayes Approach to Statistics.” In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, 157–163. University of California Press. [25](#), [33](#)
- [135] — (1964). “The Empirical Bayes Approach to Statistical Decision Problems.” *Annals of Mathematical Statistics*, 35: 1–20. [25](#)
- [136] Ross, I. and Tukey, J. W. (1975). *Index to Statistics and Probability: Permuted Titles, A—Microbiology*. Los Altos, CA: R&D Press. [15](#)
- [137] Rubin, H. (1949). “The Existence of Measurable Utility and Psychological Probability.” *Statistics* 331, Cowles Commission Discussion Paper. 5 pp. [13](#)
- [138] Savage, L. J. (1951). “The Theory of Statistical Decision.” *Journal of the American Statistical Association*, 46: 55–67. Reprinted in Savage, Leonard J. (1981). *The Writings of Leonard Jimmie Savage: A Memorial Selection*. American Statistical Association and Institute of Mathematical Statistics, Washington, DC, 201-213. [15](#), [16](#)
- [139] — (1952). “Review of *Logical Foundations of Probability* by Rudolph Carnap.” *Econometrica*, 29: 688–1690. [21](#)
- [140] — (1954). *The Foundations of Statistics*. New York: Wiley, second revised paperback edition, dover, new york, 1972 edition. [9](#), [16](#), [18](#)

- [141] — (1960). “Unpublished Reading Note.” Reproduced in Dale (36) as an Appendix, pp. 597-602. 3
- [142] — (1961). “The Foundations of Statistical Inference Reconsidered.” In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, 575–586. 23, 28
- [143] — (1962). *The Foundations of Statistical Inference. A Discussion*. London: Methuen. G. Barnard and D.R. Cox, eds. 17, 19
- [144] — (1970). “Reading Suggestions for the Foundations of Statistics.” *American Statistician*, 24: 23–27. [Reprinted in Savage, Leonard J. (1981). *The Writings of Leonard Jimmie Savage: A Memorial Selection*. American Statistical Association and Institute of Mathematical Statistics, Washington, DC, 536-546.]. 16
- [145] — (1976). “On Rereading R.A. Fisher (with discussion) (J.W. Pratt, ed.)” *Annals of Statistics*, 4: 441–500. 25
- [146] — (1981). *The Writings of Leonard Jimmie Savage: A Memorial Selection*. Washington, DC: American Statistical Association and Institute of Mathematical Statistics. 28
- [147] Schlaifer, R. (1959). *Probability and Statistics for Business Decisions*. New York: McGraw-Hill. 18
- [148] Scott, A. and Smith, T. (1969). “Estimation for Multi-Stage Surveys.” *Journal of the American Statistical Association*, 64: 830–840. 25
- [149] Seidenfeld, T. (1992). “R.A. Fisher’s Fiducial Argument and Bayes’ Theorem.” *Statistical Science*, 7: 358–368. 23
- [150] Seneta, E. (1998). “Early Influences on Probability and Statistics in the Russian Empire.” *Archive for History of Exact Sciences*, 53: 201–213. 6
- [151] Shafer, G. and Vovk, V. (2003). “The Sources of Kolmogorov’s *Grundbegriffe*.” Working Paper #4. Available at <http://www.probabilityandfinance.com>. 6
- [152] Sheynin, O. B. (1989). “A. A. Markov’s Work on Probability.” *Archive for History of Exact Sciences*, 39: 337–377. [Errata in 40 (1989), 387.]. 6
- [153] — (2004). “Fechner as a Statistician.” *British Journal of Mathematical and Statistical Psychology*, 57: 53–72. 6
- [154] Smith, A. F. M. (1995). “A Conversation with Dennis Lindley.” *Statistical Science*, 10: 305–319. 11
- [155] Stigler, S. M. (1978). “Francis Ysidro Edgeworth, Statistician.” *Journal of the Royal Statistical Society, Series A*, 141: 287–322. (Also available with some corrections as Chapter 5 of Stigler, Stephen M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge, MA.). 7

- [156] — (1980). “Stigler’s Law of Eponymy.” *Transactions of the New York Academy of Sciences*, Ser. 2, 39: 147–158. (Also available as Chapter 14 of Stigler, Stephen M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge, MA.). 2
- [157] — (1982). “Thomas Bayes’ Bayesian Inference.” *Journal of the Royal Statistical Society, Series A*, 145: 250–258. 4
- [158] — (1983). “Who Discovered Bayes’ Theorem?” *American Statistician*, 37: 290–296. (Also available with some corrections as Chapter 15 of Stigler, Stephen M. (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Harvard University Press, Cambridge, MA.). 2, 4
- [159] — (1986). *The History of Statistics: The Measurement of Uncertainty Before 1900*. Cambridge, MA: Harvard University Press. 2
- [160] — (1986). “Laplace’s 1774 Memoir on Inverse Probability.” *Statistical Science*, 1: 359–363. 5
- [161] — (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press. 2
- [162] — (2004). “How Ronald Fisher Became a Mathematical Statistician.” Presented at *Journée “Bernard Bru”*, Université de Paris VI, April 2, 2004. 8
- [163] Student (1908). “The Probable Error of a Mean.” *Biometrika*, 6: 1–25. 7
- [164] Suppes, P. (1956). “The Role of Subjective Probability and Utility in Decision-making.” In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 5, 61–73. 22
- [165] Tiao, G. and Zellner, A. (1964). “Bayes’ Theorem and the Use of Prior Knowledge in Regression Analysis.” *Biometrika*, 51: 219–230. 26
- [166] Tukey, J. W. (1990). “The Collected Works of John W. Tukey, Volume VI: More Mathematical, 1938 - 1984.” In Mallows, C. L. (ed.), *The Collected Works of John W. Tukey, Volume VI: More Mathematical, 1938 - 1984*, 119–148. Pacific Grove, CA: Wadsworth. 22
- [167] — (1990). “The Present State of Fiducial Probability.” In Mallows, C. L. (ed.), *The Collected Works of John W. Tukey, Volume VI: More Mathematical, 1938 - 1984*, volume VI, 55–118. Pacific Grove, CA: Wadsworth. 22
- [168] von Mises, R. (1928). *Wahrscheinlichkeitsrechnung, Statistik und Wahrheit*. Dover. 8
- [169] — (1941). “On the Foundations of Probability and Statistics.” *Annals of Mathematical Statistics*, 12: 191–205. 8
- [170] — (1942). “On the Correct Use of Bayes’ Formula.” *Annals of Mathematical Statistics*, 13: 156–165. 8

- [171] von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press. (Paperback edition, 1980.). 16
- [172] Wald, A. (1939). “Contributions to the Theory of Statistical Estimation and Testing – A Hypotheses.” *Annals of Mathematical Statistics*, 10: 299–326. 9
- [173] — (1947). “An Essentially Complete Class of Admissible Decision Functions.” *Annals of Mathematical Statistics*, 18: 549–555. 9
- [174] — (1947). *Sequential Analysis*. New York: Wiley. (Paperback edition, Dover, New York, 1973.). 13
- [175] — (1950). *Statistical Decision Functions*. New York: Wiley. 15
- [176] Wallis, W. A. (1980). “The Statistical Research Group, 1942–1945.” *Journal of the American Statistical Association*, 75: 320–335. 11, 13
- [177] — (1981). “Memorial Tribute.” In *Savage, Leonard J. The Writings of Leonard Jimmie Savage: A Memorial Selection*, 11–24. Washington, DC: American Statistical Association and Institute of Mathematical Statistics. 17
- [178] Wilks, S. S. (1936). “Discussion of Ernest Nagel: ‘The Meaning of Probability.’” *Journal of the American Statistical Association*, 31: 29–30. 9
- [179] Wrinch, D. and Jeffreys, H. (1919). “On Some Aspects of the Theory of Probability.” *Philosophical Magazine*, 38: 715–731. 10
- [180] Zabell, S. (1982). “W.E. Johnson’s ‘Sufficientness’ Postulate.” *Annals of Statistics*, 10: 1090–1099. 10
- [181] — (1989). “R.A. Fisher on the History of Inverse Probability.” *Statistical Science*, 4: 247–256. 6
- [182] — (1989). “R.A. Fisher on the History of Inverse Probability: Rejoinder.” *Statistical Science*, 4: 261–263. 6
- [183] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley. 26

Acknowledgments

I am indebted to many people for input to the enterprise behind this paper, for informal conversations, the sharing of papers and other materials. In particular, Margo Anderson sparked my interest in the question behind the paper, and Steve Stigler has made several suggestions, helped me track down information on the University of Chicago in the 1950s, and provided me with repeated edits. Teddy Seidenfeld also provided considerable assistance and suggestions. I have been especially fortunate to have had specific comments, recollections and/or other input from Robert Aumann, David Brillinger, Art Dempster, I.J. Good, Joel Greenhouse, Wes Johnson, Dennis Lindley, Duncan Luce, Al Madansky, Ward Edwards, Frederick Mosteller,

John Pratt, Howard Raiffa, Herman Rubin, Alastair Scott, Patrick Suppes, and Arnold Zellner, although none are responsible for the interpretations or the arguments in this paper. A presentation of much of the contents of the paper at the Paris Seminar on “Histoire du calcul des probabilités et de la statistique,” also produced a number of helpful suggestions and comments, especially from Marc Barbut and Antoine de Falguerolles. Finally, I thank the editor and referees for providing detailed comments and helping me to address both language and historical issues. This preparation of the paper was carried out while I was a visiting researcher at the Centre de Recherche en Economie et Statistique of the Institut National de la Statistique et des Études Économiques, Paris, France.

Corrected May 16, 2007.