Arch. Hist. Exact Sci. (2009) 63:51–79
DOI 10.1007/s00407-008-0036-8

Burnside's engagement with the "modern theory of statistics"

John Aldrich

Received: 25 March 2008 / Published online: 8 November 2008

Abstract The group theorist William Burnside devoted much of the last decade of his life to probability and statistics. The work led to contact with Ronald Fisher who was on his way to becoming the leading statistician of the age and with Karl Pearson, the man Fisher supplanted. Burnside corresponded with Fisher for nearly three years until their correspondence ended abruptly. This paper examines Burnside's interactions with the statisticians and looks more generally at his work in probability and statistics.

1 Introduction

In April 1919, 2 months short of his 67th birthday, William Burnside retired from the Royal Naval College at Greenwich. In retirement the renowned group theorist spent much of his time on probability, producing ten short papers and the manuscript for a book—perhaps half of his output in those years. The first fruit of the probability project—a paper on the theory of errors (1923b)—led to contact with Ronald Fisher. A correspondence of nearly 3 years followed in which Burnside wrote 25 letters, some much longer than the published papers. In the second letter he told Fisher, "I have no proper acquaintance with either the phraseology or the ideas of the modern theory

Communicated by J.J. Gray.

I am grateful to June Barrow-Green and Tony Mann for showing me important letters, to Susan Woodburn, Elise Bennetto and Jeniffer Beauchamps of the University of Adelaide for help with letters to Fisher and to Jonathan Harrison of St. John's Cambridge for help with letters to Baker and White. Peter Neumann, the editor and a referee made helpful comments on an earlier draft.

J. Aldrich (✉)
Economics Division, School of Social Sciences, University of Southampton,
Southampton SO17 1BJ, UK
e-mail: john.aldrich@soton.ac.uk

of statistics." With Fisher providing reading, Burnside was soon acquainted enough to criticise the work of Karl Pearson and then of Fisher himself. Burnside chose his statistical targets well—Fisher the coming man and Pearson the come—though neither was one to admit error. Burnside's engagement with the statisticians may seem surprising in one who "detested" mathematical controversy (Forsyth 1928a, p. xvii) for theirs was a field where controversy ruled and his interventions only invited more!

The appearance of an edition of his *Collected Papers*—Neumann et al. (2004)—makes this a good moment to review Burnside's work in probability and statistics. The edition reproduces the papers—bar one—and contains background chapters on the life and work but probability and statistics intrude into these only when Everett et al. (2004, p. 104), report Burnside being upset by a letter from Fisher—see Sect. 8. Individual pieces by Burnside have been noticed by historians of statistics—including Edwards (1978), Dale (1991) and Pfanzagl and Sheynin (1996)—but not as part of a larger effort. Nothing has been published about Burnside's criticism of Fisher's work or about his relationship with Fisher: the "Note on Dr. Burnside's recent paper on errors of observation" is listed with Fisher's other writings in Box's (1978) biography but Burnside is not mentioned in the text.

Burnside's work in probability and statistics did *not* make a difference to those subjects. His probability was behind continental work and, while Fisher and Pearson responded in print to his criticisms, they were not deflected by them. Burnside's activity in probability and statistics is interesting mainly for what it shows about the state of these subjects at the time and the personalities of those involved. The letters to Fisher are the only substantial series of Burnside's letters extant. Fisher's letters are lost which is a shame as the exchanges were largely about *his* theory of statistics and what it rested on, matters he seems to have discussed only with Burnside. Among letters to Fisher from this period—one of the most productive of his life—the Burnside letters complement the 50 from W. S. Gosset (McMullen 1970) and the handful from Leonard Darwin (Bennett 1983). With each correspondent Fisher discussed different things and had a different relationship—Burnside was the mathematical correspondent and the troublesome one.

In the account below Sect. 2 has some background on persons and subjects, Sects. 3 and 4 describe Burnside's papers on the theory of errors and how the statisticians reacted to them, Sects. 5–8 treat Burnside's criticisms of Pearson and Fisher, Sect. 9 considers his non-statistical probability work and the posthumously published *Theory of Probability* (1928) and Sect. 10 has some concluding remarks.

2 Cambridge mathematicians and probability

William Burnside (1852–1927), Karl Pearson (1856–1936) and Ronald Fisher (1890-1962) were Cambridge mathematicians all successful enough to become fellows of the Royal Society: Burnside was second wrangler in 1874 and FRS in 1893, Pearson third wrangler in 1879 and FRS in 1896 and Fisher a wrangler in 1912 (by then the order of merit had gone) and FRS in 1929. All made their careers away from Cambridge and their lasting reputations away from the applied mathematics in which they had been trained. From 1885 until he retired Burnside was professor of mathematics at Greenwich. He taught applied mathematics to naval officers but his distinction as a mathematician rested on his research in group theory on which he began publishing in 1893. Pearson was appointed professor of applied mathematics at University College London in 1884 and taught that subject for nearly thirty years although from the mid-1990s biometry was his main interest; see Stigler (1986, Part 3) for Pearson's early work and the background to it. Fisher moved furthest and fastest from applied mathematics: he taught himself genetics and biometry as an undergraduate and from 1919 was the statistician at Rothamsted experimental (agricultural) station. For further biographical information see Forsyth (1928a), Everett et al. (2004) and Neumann et al. (2004) for Burnside, Pearson (1936/8) for Pearson and Box (1978) for Fisher.

Pearson knew Burnside as a teacher—see Pearson (1936a, p. 29)—and until around 1900 they belonged to the same mathematical community but by 1923 Burnside inhabited one world and Pearson and Fisher another, with journals and Cambridge providing only very weak links. Pearson dropped his connections with Cambridge but Burnside kept his and Fisher was re-establishing his as a non-resident fellow of Caius College. Burnside was a frequent contributor to the *Proceedings of the Cambridge Philosophical Society*, the house journal of the Cambridge mathematicians; he also published in the *Messenger of Mathematics* (another Cambridge-edited journal) to which Fisher had contributed in 1912 and 1913. The *Philosophical Magazine* was known to all and Burnside and Pearson published there occasionally; the journals of the Royal Society were also known—they had been vital for Pearson before he started *Biometrika* in 1901 and Fisher had a major paper in the *Philosophical Transactions* in 1922. That paper, "On the Mathematical Foundations of Theoretical Statistics," (the "Foundations" for short) is central to the present story and indeed to the history of mathematical statistics—"arguably the most influential paper on that subject in the twentieth century" in (Stigler, 2005, p. 32) judgement.

Pearson and Fisher were in the same world but they found coexisting very difficult: see Box (1978, ch. 3). When Pearson refereed Fisher's big paper on population genetics for the Royal Society he was cool and it appeared elsewhere as Fisher (1918)—see Norton and Pearson (1976). In 1920 Pearson rejected a paper Fisher had sent to *Biometrika*—this appeared elsewhere as Fisher (1921); see Pearson (1968, p. 453) and Bennett (1983, pp. 73–4). *Biometrika* had turned away Fisher's work twice before and he sent no more. If Pearson was the Pope, Fisher was out of communion—if not a heretic.

In 1920 when Burnside was contemplating a book on probability there were no modern books in English, or specialists like Czuber or Markov to write them: Irwin (1967, p. 151) recalled that when he started lecturing in 1921 he used Poincaré (1912), Whitworth (1901), and Venn (1888). Yet from the time of Maxwell British applied mathematicians had been turning their hand to probability and Burnside, Pearson and Fisher were among those who kept turning—see Sect. 9. Burnside's experience in probability was limited but it went back a long way—possibly to his school-days and textbooks like Todhunter's *Algebra* (1858) which used probability to illustrate combinatorial methods. He may have well have studied probability and the theory of errors as an undergraduate—5 years later Pearson did. Burnside read the proofs and made the index for the second, 1883, edition of Thomson and Tait's standard work; its chapter III on "experience" has a brief treatment of the theory of errors. In 1888–89 Burnside published on the kinetic theory of gases—Barrow-Green (2004) reviews all his contributions to applied mathematics—and his articles (1888, 1889) with their complicated calculations of averages, involve probability even though the term never appears. French military schools taught the theory of errors—see Bru (1996) on the teaching in Broemeling and Broemeling (2003) on the work of the artillery captain Ernest Lhoste—but it seems not to have figured in Burnside's teaching at Greenwich. Forsyth (1928a, p. xiii) describes his two elementary courses as concerned with "principles of ballistics" and with "strength of materials, dynamics and heat engines" and the most advanced with "kinematics, kinetics and hydrodynamics."

Burnside seems to have conceived the probability project after writing "On the Probable Regularity of a Random Distribution of Points" (1919), a paper Forsyth (1928b, p. v) describes as treating what is "manifestly a military question, reduced (for purposes of calculation) to a purely mathematical form." Forsyth (1928a, p. xxix) suggests that Burnside's interest in probability came about as a result of war work; his efforts to get such work are recounted by Everett et al. (2004, p. 105) but how—or whether—this paper fits in is not known. The probability project did not go to plan for the first installment—a paper on errors of observation—provoked the appearance of the "modern theory of statistics." The paper and its unforeseen consequences will occupy us until Sect. 9 when we return to the larger project.

3 On errors of observation 1923

"Many high authorities have asserted that the reasoning employed by Laplace, Gauss and others is not well founded" wrote Thomson and Tait (1867, p. 312) and fifty years later the warning (or challenge) was still in force with no authority to suppress the grumbling or impose a reform: "it would be difficult to find a subject where confusion of thought and loose reasoning are more prevalent" lamented the astronomer Stewart (1920, p. 218). The subject was open to all-comers and their improvements appeared in mathematics, astronomy and general science journals. The 'contributions' generally failed to register and were no more cumulative than solutions to newspaper puzzles—indeed Stewart's improvement had been published by Pearson (1901) in the same journal twenty years before. Probability—see Sect. 9 below—was just as open and unprogressive but there was no comparable standing challenge unless it was the much bigger question of the nature of probability itself.

Burnside's "On errors of observation" in the *Proceedings of the Cambridge Philosophical Society* was a typical product of this state of affairs. Aiming to correct an established conception, it (1923b, p. 482) refers only to what is "standard." Given observations $x_1, \ldots, x_n$, the "result is generally stated as follows …the true value of the constant [x_0] is equally likely to lie within or without the range from $\overline{x} - \epsilon$ to $\overline{x} + \epsilon$ ", where the probable error ϵ is given by

$$\epsilon = 0.6745\sqrt{\frac{\sum_1^n (x_i - \overline{x})^2}{n(n-1)}}$$

The value 0.6745 is the 25th percentile of the standard normal distribution. In modern classical statistics the formula for $\overline{x} \pm \epsilon$ would be interpreted as that for a 50% large sample confidence interval for x_0.

The work of Student (W. S. Gosset) is central to modern treatments of the problem and has been so since Fisher put it there in his Burnside years. Student’s *Biometrika* paper began with a statement of what is “well known” and from his own resources—with a little help from Pearson—improved upon it. Otherwise the paper was not typical. It was far out of the theory of errors mainstream and—with examples of small samples from medicine and agriculture—far even from the biometric stream. It did more than register, it achieved canonisation for publication in *Biometrika* and inclusion in Pearson’s (1914) tables signified as much for the biometricians.

The “well known” result of Student (1908, p. 2) is that “when the sample is large” the quantity

$$z = \frac{\overline{x} - x_0}{\sqrt{\frac{\sum_1^n (x_i - \overline{x})^2}{n}}}$$

is normally distributed with mean 0 and variance n. The difficulty was that “no one has yet told us very clearly where the limit between “large” and “small” is to be drawn.” The denominator in z is not constant across samples and Gosset realised that the assumption that it is leads to serious errors when n is small. He obtained—or rather conjectured correctly—the distribution of z. Student’s solution was transformed into the t distribution with $n - 1$ degrees of freedom when Fisher (1924/8) described a family of distributions based on the normal distribution and in the process replaced z with $t = z\sqrt{n-1}$. These developments have generated a considerable historical literature, ranging from Eisenhart (1979) on the transition of z to t to Aldrich (2003) on Student’s role in the development of what Burnside called the “phraseology” of modern statistics.

Where Student worried about the applicability of the usual result, Burnside (1923b, p. 486) worried about the way it was derived or the way he thought it was derived:

> The standard formula is arrived at by assuming (i) that the errors follow Gauss’s law, (ii) that the precision-constant has that particular value which makes the probability of the observed set of values as great as possible. The alternative formula here put forward ... is arrived at by making the same first assumption and replacing the second by the assumption that à priori all values of the precision-constant are equally likely.

Burnside writes Gauss’s law with precision-constant h (corresponding to $\frac{1}{\sigma\sqrt{2}}$ of the statisticians’ normal distribution) as

$$\sqrt{\frac{h}{\pi}} e^{-hy^2} \delta y$$

which gives “the probability of an error in one determination lying between y and $y + \delta y$.” “The à priori probability that the precision-constant lay between h and $h + \delta h$” is $p\delta h$; the “simplest” assumption to make is that p is independent of h, “i.e. that à priori all values of the precision constant are equally likely.”

Burnside strives for the impossible, a posterior probability statement (the 50% interval) about x_0 having specified only the distribution of the errors and a prior for h (later he acknowledged the point in a letter to Fisher—see Sect. 4 below). In a survey of early Bayesian work leading to the t form, Pfanzagl and Sheynin (1996, p. 893) relate their experience of reading Burnside, "Regrettably the authors have been unable to obtain further insights from the proofs since these are hard to follow, even after correcting a misleading misprint .." The other authors they consider—L üroth, Edgeworth and Jeffreys—extended Gauss's first derivation of least squares to accommodate unknown precision by specifying a joint prior for location and precision, obtaining the joint posterior and then integrating out the precision to obtain a marginal posterior for location.

Burnside focusses on the quantity ϵ and his idea is that behind the accepted formula is the proposition that the estimation error $(\overline{x} - x_0)$ is normally distributed with mean zero and variance s^2/n and that this proposition follows from the unreasonable assumption that the precision-constant is known to coincide with its estimated value. Instead of making that assumption he uses the posterior distribution of the precision-constant based on a uniform prior for it and forms a mixture of the conditional normals given the precision-constant. Forsyth (1928a, p. xxix) hints that behind the paper was some practical interest but the result is a very academic piece: it compares the new intervals (for sample sizes up to 10) with the conventional one without noting that the former are narrower and finishes lamely with, "It is a matter of individual judgment which form of [assumption about h] is the more reasonable," without indicating the factors on which the judgment might be based.

Burnside's paper registered and produced responses in the *PCPS* and in *Biometrika*—independent responses with only a reference to Student in common. Fisher's note in the *PCPS*—described in the next section—was designed to lead to greater things but the note in *Biometrika* aimed simply to cancel Burnside's contribution and preserve the status quo. Ethel Newbold wrote at the prompting of Major Greenwood who had noticed that Burnside's formula gave a narrower interval than the conventional normal one; Newbold, who worked also with Pearson, died young in 1933 and is remembered by Greenwood (1933). Her paper considered the relationship between Burnside's distribution and Student's from different points of view. She reproduced Burnside's argument with other measures of precision in place of Burnside's h, although she (p. 402) doubted whether an observer could ever be "in complete ignorance of the precision of his observations." She showed how a uniform prior for the parameter S where $S^m = h$ (for some m) generates a t-density with a different number of degrees of freedom (this term was introduced later by Fisher (1924/8)) according to the choice of m. Student's case is shown to correspond to $1/m = 0$. Newbold was not much impressed with the general scheme she had devised for, crucially, the Student case fits "without burdening itself with the à priori assumptions attached to the other cases." (p. 405). She repeated Student's sampling experiments and found—not surprisingly—that Burnside's distribution gave a much worse fit to the empirical distribution. She (p. 406) amended Burnside's conclusion to "very little scope is left for individual judgement in choosing between Students' formula and that of Dr. Burnside." For *Biometrika* the subject was closed. Burnside probably never saw Newbold's note. The note was relevant to Jeffreys's (1931, pp. 66ff) discussion of the representation of ignorance about precision but, like Burnside, Jeffreys did not expect enlightenment from this quarter; it is more surprising that he missed Burnside who was closer to home—see Aldrich (2005b).

4 Statistics is announced and a correction made 1923

Fisher's note in the *PCPS* took a different line from the official one in *Biometrika*. It opens magisterially:

> That branch of applied mathematics which is now known as Statistics has been gradually built up to meet very different needs among different classes of workers. Widely different notations have been employed to represent the same relations, and still more widely different methods of treatment have been designed for essentially the same statistical problem. It is therefore not surprising that Dr. Burnside writing on errors of observation in 1923 should have overlooked the brilliant work of "Student" in 1908, which largely anticipates his conclusion (1923, p. 635).

In 1923, the subject depicted in the first sentence existed mainly in Fisher's imagination; there was a subject, "statistics," dominated by economists and given institutional form in the Royal Statistical Society—all this is described in Aldrich (2008)—but that was not what Fisher had in mind. He was pressing for the establishment of a chair of mathematical statistics in Cambridge (Box (p. 239)) and the declaration that statistics is a branch of applied mathematics was of a piece with this. It is implicit in this declaration that "Statistics" has sovereignty over the theory of errors and implicit in the note as a whole that the biometricians have no precedence among the "different classes of workers." Most immediately and practically he was creating an opening for statistics and himself and five more papers of his would appear in the *PCPS*.

Unlike Newbold, Fisher did not criticise Burnside, though he thought even less of Burnside's Bayesian assumptions, having exposed their "baseless character" in the "Foundations" (p. 326): he only observes that the difference in the formulae "is traceable to Dr. Burnside's assumption of an à priori probability for the precision constant, whereas Student's formula gives the actual distribution of z in random samples." Fisher was more severe in 1941 when he (pp. 141–2) used Burnside's paper to illustrate all that had been wrong with the theory of errors before Student. In 1923 his object was to expose the mathematicians to Statistics and to Gosset's work which "deserves to be far more widely known than it is at present." He was also exposing his own work for the derivation he gave was the one he had devised a decade before but which had never been published. In 1912, Fisher sent the derivation to Gosset, who forwarded it to Karl Pearson for consideration in *Biometrika*; see Pearson (1968, p. 446). Pearson did not publish the derivation, although Fisher used parts of it in other publications. The note, received by the *Proceedings* on July 16th, belonged to the past in the further sense that Fisher (1922b) had already extended Student's z to regression—see Aldrich (2005a) for a review of Fisher's regression work—and was working towards the system of distributions in which z would be replaced by t; see his letter to Gosset of May 2nd reproduced in McMullen (1970) and Box (p. 118). In July, Fisher sent

Gosset Burnside's paper and his own note; Gosset merely commented, "It is interesting to see how à priori probability has got him just off the line." (McMullen 1970).

By August, Burnside and Fisher were corresponding—they never met. Burnside had a "fierce but short-lived temper" and so had Fisher (Everett et al. 2004, p. 105; cf. Yates and Mather 1963, p. 97) but they managed to contain their tempers for nearly three years. In the beginning Burnside wanted to learn about Student and about the work of statisticians generally: in his second letter he wrote, "I have no proper acquaintance with either the phraseology or the ideas of the modern theory of statistics." Burnside did not want to impose on the busy younger man but Fisher seemed prepared to give a correspondence course on the "Foundations.". Generally Burnside would ask a question and Fisher's reply would set off an exchange of letters. Having only Burnside's letters is a great handicap for they show such misunderstanding that inferences from them to what Fisher was saying are quite precarious.

Beginning at the beginning in August 1923 with the theory of errors there were letters to Fisher on the 4th, 8th, 11th and 20th; bursts of letters separated by months of silence would be the pattern. Fisher must have written first as Burnside's first letter is a reply. He begins by saying he would not have published had he known of Student's paper and ends by asking, as everybody did. "Do you happen to know who "Student" is?" Burnside admitted, "I seldom go to London and so still have not seen Student's paper" and he probably thought the proof in Fisher's note was a paraphrase of Student's; so his comment on the "really brilliant point of Student's work" was actually in praise of Fisher. Burnside explained that he had "had on hand" for "two or three years" a book on the "Calculation of Probabilities" and that he proposed to include "Student's result" and he wanted to try his proof on Fisher. The proof he sketched was of the Student-Fisher result *not* of the Bayesian result he gave earlier in the year. It involved an orthogonal transformation of the observations and presented the joint density of $\sum x_i$ and $\sum(x_i - \overline{x})^2$. It was an analytic counterpart to the geometrical proof given by Fisher. Unknown to them—and to the British statistical community before Pearson (1931) drew attention to it—the geodesist Helmert (1876) had used the same transformation; see Sect. 9. In his letter, Burnside made a point of saying that if $n = 3$, "a geometrical conception of this step is possible" but if $n > 3$, the result "can only be proved by an analytical process." Naturally Fisher disagreed as he had attacked the general case geometrically; Burnside's next letter begins, "I think we must just agree to differ as to what a geometrical proof is." When Fisher next wrote about Student's result applied to regression coefficients in the "Applications of Students' distribution" he may have had Burnside in mind:

> It is perhaps worthwhile to give, at length, an algebraical method of proof since analogous cases have hitherto been demonstrated only geometrically, by means of a construction in Euclidian hyperspace, and the validity of such methods of proof may not be universally admitted (1925c, p. 97).

Fisher's orthogonal transformation specialised to the normal mean case is the one Burnside proposed in August 1923 and which he published in the *Theory of Probability.* The derivation, or rather the Helmert transformation on which it was based was good enough to be re-rediscovered: Irwin (1931) did so and then (1934, p. 242) discovered Burnside's priority. See Sect. 9 below.

Now that Burnside was reproducing Student–Fisher he reconsidered his original argument to surprising effect. On October 20, he informed Fisher he was sending a note to the *PCPS* for "I have now discovered what I consider a gross error at the end of that paper" and he described "what I really proved (and I believe correctly proved from the premisses assumed)." This was Student's result! Burnside finished the letter by declaring, "Personally I believe that [any probability statement] about the constant could only be proved when some assumption had been made with respect the probability of the constant ... *before* any determinations had been carried out." This remarkable letter reflects a very strange view of what Fisher stood for.

The published correction note (1923c) begins, "The object of this note is to direct attention to an entirely false inference which was drawn from the analysis given in my previous paper ..." The "false inference" involved going, without justification, from the sampling distribution of the quantity

$$\frac{\overline{x} - x_0}{s}, \quad \text{where } s^2 = \frac{\sum_1^n (x_i - \overline{x})^2}{n},$$

to the Bayes posterior for x_0. The false inference professed "to determine, after a set of n observations, the probabilities of various values x_0 without any assumption regarding their probabilities before the observations." Burnside pointed out correctly that he had introduced no prior for x_0 and that in his analysis he had treated x_0 as a constant and had integrated the errors over an infinite range. However, he did not look in detail at his argument or explain what the prior for h was doing there.

The correction note was extraordinary. It adopted the Student–Fisher view of the problem—including some of their notation—without mentioning them and then claimed that the original paper contained a derivation of their result and was marred only by an inference too far. One can only guess what Fisher thought of this; he could not have exploded for Burnside's letters carried on with no reference to the development, let alone explanation. In his memoir Forsyth (1928a, p. xxv) wrote that Burnside "did not hesitate to abandon an opinion, if ever he regarded it as not fully tenable."

5 Struggling with the "Foundations" 1923–1926

As well as worrying about the derivation to go into his book—see Sect. 9 below for the conclusion to that story—Burnside wanted to understand the theory of statistics, which meant understanding the "Foundations." He may well have been its first real reader—the first to try to understand it line by line. Bowley offered Fisher some "rather hasty remarks" in a letter of 25 May 1921 and the referees, Udny Yule and Arthur Eddington, aware of Fisher's quality from his other work seem only to have sniffed at the paper before approving it (their reports are in Edwards (1997, p. 181)). No established authors paid any attention to the "Foundations" and even their criticisms appeared only much later, see Bowley (1935) or Pearson (1936b). Fisher's ideas were taken up by new people—Neyman and E. S. Pearson in 1928 and Hotelling (1930);

their's was an easier path than Burnside's because in the interim Fisher produced more accessible works and Hotelling even had the advantage of a stay at Rothamsted.

Burnside had been reading the "Foundations" even before Fisher contacted him and in the third letter (August 11th) he had detailed questions as well as some comments on Fisher (1915a,b)—his most ambitious geometrical venture to date—which Fisher must have sent. The list of neologisms that opens the "Foundations" might have warned Burnside that this was not going to be an ordinary paper; for an account of it see Aldrich (1997) and Stigler (2005). However, Burnside thought it could be approached in the normal way and asked Fisher for a reference where the underlying ideas are "more or less systematically expounded." Fisher put him on to Yule's (1911) *Introduction* which he found more useful than Soper's (1922) *Frequency Arrays*, although there was still a great gap between Yule's book and Fisher's article.

In 1923 Burnside made a serious effort to understand the "Foundations", or rather its Sect. 6 on "formal solution of problems of estimation" which he thought bore most closely on topics in his book—Bayes' theorem and the theory of errors. At first viewing he was not persuaded by Fisher's method of maximum likelihood and he could not follow the argument on the large sample distribution of the maximum likelihood value. He was never interested in Fisher's larger vision—information and sufficiency—not to speak of the statistical or biometric applications that Fisher was interested in. Burnside kept coming back to the "Foundations" although in the end (see Sects. 7 and 8 below) his interest was entirely concentrated on two sentences about the "infinite hypothetical population." After lapsing in October 1923 the correspondence resumed at the end of 1924 although there was a short exchange in mid-1924 after Burnside had challenged Karl Pearson over Bayes' theorem.

6 The fundamental problem of practical statistics 1924

Statistics had come to Burnside but he now took the initiative. His move against Karl Pearson was sharp and clean: he saw an error and exposed it. However he went away frustrated when Pearson countered that the misunderstanding was all his. *Biometrika* was off the beaten track for Burnside and his note there is missing from the *Collected Papers* although it is known to historians of statistics interested in Pearson or in Bayes: it has been discussed by Edwards (1978, p. 118), Dale (1991, pp. 377–392), Hald (1998, pp. 253–6) and Aldrich (2007). These papers detail the Pearson background but for Burnside there was no background, only the paper and the error it contained.

In the correction note Burnside (1923c, p. 27) had reservations about the use of the Bayesian argument in the case of no prior information:

> Many opinions seem to be held at the present time with regard to Bayes's formula for the probability of causes. I have, however seen no support for the supposition that it is possible, after an observed event, to determine the respective probabilities of its various causes, in the case in which nothing is known about the corresponding probabilities before the event.

In November he had told H. F. Baker, friend as well as editor of the *PCPS*, "You will wonder why I have not yet sent you [the correction note] ... I can't quite make up my mind as to how far I ought to refer to Bayes' theorem. Some such reference is absolutely necessary: but it involves dealing with a point on which there seems to be great difference of opinion."

In the "Foundations" Burnside would have found three opinions: Fisher's (Sect. 6, p. 324) own, that the use of a uniform prior in the case when nothing is known about the value of the parameter ("Bayes' postulate") is inappropriate, plus two more he invoked to show how the basic principles of statistical science "are still in a state of obscurity":

> [in Pearson (1920)] one of the most eminent of modern statisticians presents what purports to be a general proof of Bayes' postulate, a proof of which in the opinion of a second statistician of equal eminence "seems to rest upon a very peculiar—not to say hardly supposable—relation." (Sect. 1, p. 310)

Burnside tracked down Pearson's grandly titled piece, "The fundamental problem of practical statistics," but not the work of the second statistician Edgeworth (1921).

The "fundamental problem" of Pearson (1920, p. 1) is:

> An "event" has occurred p times out of $p + q = n$ trials, where we have no a priori knowledge of the frequency of the event in the total number of occurrences. What is the probability of its occurring r times in a further $r + s = m$ trials?

This is the prediction variant (due to Laplace) of Bayes's (1763, p. 376) problem in the doctrine of chances: "*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." The conventional solution to the fundamental problem is that the required probability is

$$\frac{B(p+r+1, q+s+1)}{B(p+1, q+1)B(r+1, s+1)},$$

a result reached by a standard argument assuming Bernoulli trials and a uniform prior for the probability of an occurrence—the "equal distribution of ignorance" or "Bayes' postulate."

The 1920 paper is a marvellous Pearsonian production with a lively historical survey and a demonstration that "the fundamental result of Laplace in no way depends upon the equal distribution of ignorance" (pp. 5–6) leading into a long discussion of the evaluation of beta functions. The demonstration was astounding, especially coming after decades of debate—usually around the "rule of succession" case ($q = 0$, $m = 1$)—when all, including Pearson (1892, 1907), had taken it for granted that the result depends on the equal distribution of ignorance. As recently as 1917, Pearson—in Soper et al. (1917, p. 353)—had chastised Fisher for making this assumption in a correlation version of Bayes's original problem.

Pearson's new thinking was based on a re-examination of Bayes's essay. Bayes had used a physical model when making his argument but the model did not figure in later contributions: it is described by Todhunter (1865, pp. 294–300) but as *history* for the statement of the problem involved no reference to it and Todhunter was clear that Bayes's own way of reaching his conclusion had been superseded. Stigler (1982, p. 255) reports how Pearson resurrected the model partly from a growing interest in the history of statistics and partly because he used something like it in his own work. The feature common to Pearson and Bayes is that an event occurs when an underlying continuous variable passes a threshold value: Pearson imagines "men sickening from a disease when their resistance falls below a certain level" (1921, p. 301), Bayes imagines balls rolled on a table with the threshold defined by the stopping point of the first ball, so that an *occurrence* is registered when a subsequent ball stops beyond the first. In 1920, Pearson argued that Bayes's assumption of a uniform distribution for the threshold value could be replaced with any continuous distribution (in his notation represented by $\phi(x)\delta x/a$) without changing the final result.

Fisher's statistician of "equal eminence" was F. Y. Edgeworth, the Oxford professor of political economy. Edgeworth was an authority but his authority in statistics was confined to a remote corner of the economic world mapped in Aldrich (2008). Edgeworth had inspired Pearson's old thinking and in a footnote to a paper in the *Statistical Journal* he (1921, pp. 82–3fn.) commented on the new:

> [it] seems to rest upon a very peculiar—not to say, hardly supposable—relation between the antecedent probability that a certain "possibility" (in Laplace's phrase) or constitution (e.g., of a coin or die) would have existed and the a posteriori probability that, if it existed, such and such events (e.g., so many Heads or Aces in n trials) would be observed. But in general if the distribution of à priori chances is $\phi(x)\delta x/a$ (Pearson, loc. cit.), the function ϕ would not reappear in the a posteriori probability.

Noting that modern terminology is not the same as Edgeworth's and that his "a posteriori probability" refers to the modern "likelihood" the objection is the same as that put by Burnside.

In the July 1921 issue of *Biometrika* Pearson replied to unnamed critics: "Some misunderstanding has arisen with regard to my paper ... I believe it is due to the critics not having read Bayes' original theorem." He (p. 301) declared further

> But if the critics say: Then this is not what we mean by Bayes' Theorem, I would reply: Quite so, but it is what Bayes meant by his own Theorem, and it probably fits much better the type of cases to which we are accustomed to apply it than what you mean by Bayes' Theorem.

This note did not settle things even at University College for Pearson (1990, p. 75) recalls that after the lectures of 1922, "Oscar Irwin and possibly others had ... concluded that K.P. had slipped up in his 1920 paper..." Burnside's note gave Pearson an excuse for returning to the subject.

Pearson published Burnside's note with a reply in the May 1924 issue of *Biometrika*; neither referred to Edgeworth's comment or to Pearson's earlier note. The one-page note "On Bayes' formula" exemplifies what Forsyth (1928a, p. xvii) calls Burnside's "efficiency of presentation." Burnside (1924c) starts with notation and a statement of "Bayes' formula" (quoted from Poincaré (1912) and essentially that given in modern textbooks). Having expressed the fundamental problem in his own notation, the note proceeds (p_{A_i} is the modern $P(A_i)$ and $p_{(A_i)B}$ is $P(B\,|A_i\,)$):

> an application of Bayes' formula is made to a case in which a change in p_{A_i} necessarily carries with it a change in $p_{(A_i)B}$; in other words this investigation is concerned with a case in which the numerical value of $p_{(A_i)B}$ depends not only on the nature of the condition A_i, but also on the probability p_{A_i} that condition A_i may be satisfied.
>
> It has been seen that, for the statistical problem stated at the beginning of the paper, the numerical value of $p_{(A_i)B}$ depends in no way on the value of p_{A_i}.
>
> There is therefore no reason for supposing that any conclusions drawn from the investigation ... will hold with respect to the statistical problem stated at the beginning of Professor Pearson's paper.

In reply Pearson (1924) took the familiar line: he begins, "Dr. Burnside, I venture to think, does not realise either the method in which I approach Bayes' Theorem, or the method in which Bayes approached it himself."

Although Burnside most probably learnt of Pearson's solution to the fundamental problem from Fisher, it figures only in one letter written after Fisher had seen the published note. In this letter of July 17th, Burnside expressed dissatisfaction with the outcome of the exchange with Pearson. Burnside had no interest in the historical Bayes—"I know nothing of Bayes' writings"—and so Pearson's appeal to history was irrelevant. In the letter Burnside was more forthright than in the note:

> It seems to me to be hardly an exaggeration to say that, in the paper in question, Pearson professes to deduce a result that would be of far reaching importance in the theory of probabilities from the fact that the value of a definite integral does not depend on the symbol used for the variable: i.e. that $\int_a^b f(x)dx$ is the same as $\int_a^b f(y)dy$.

Burnside did not pursue his disagreement with Pearson, or at least there are no letters in the Pearson archives. Burnside's intervention was probably welcome to both Pearson and Fisher but for different reasons: Pearson could reiterate his position with Burnside proof of the existence of the uninformed critic, while Fisher had at least half an ally in his campaign against Pearson and against Bayes' postulate. It is not possible to infer from Burnside's reply what Fisher had said. He probably did not say much for when he next wrote about Pearson (1920) in 1956 (!) he (1956, p. 34) repeated what he had written in the "Foundations"—that Pearson was trying to "prove" Bayes' postulate—without describing the proof or indicating that it was connected to the balls and table scheme which Fisher (p. 12) described when he presented "Bayes' analysis."

Fisher may not have replied to Burnside, or at least may not have wanted to take the discussion further, for he was busy finishing his *Statistical Methods for Research Workers* and preparing to go to Canada to present his paper (1924/8) on the distributions associated with the normal which included his reconstruction of Student's distribution. Burnside next wrote to Fisher at the end of November and there were several exchanges before Christmas 1924. It is hard to follow these exchanges—at one point there is a discussion of χ^2—but Burnside's attention was moving to what was surely the most foundational part of the "Foundations": Sect. 2 on "the purpose of statistical methods."

The "Foundations" (1922a, p. 311) describes the statistician as constructing a "hypothetical infinite population, of which the actual data are regarded as constituting a random sample." Probability is "a parameter which specifies a simple dichotomy in an infinite hypothetical [the order of the adjectives seems indifferent] population, and it represents neither more nor less than the frequency ratio which we imagine such a population to exhibit." (p. 311) Thus the meaning of the statement, "the probability of throwing a five with a die is one-sixth," is "that of a hypothetical population of an infinite number of throws, with the die in its original condition, exactly one-sixth will be fives." Burnside found it impossible to imagine a frequency ratio in an infinite population and would write two notes—(1925b) and (1926)—saying so.

7 The infinite hypothetical population 1925–1926

In March 1925 came a follow-up to the "Foundations." On the 18th H. F. Baker, the editor of the *PCPS* informed the Society's secretary, F. P. White, of the arrival of the "Theory of statistical estimation":

> This morning I receive from R. A. Fisher a paper on statistics ... I am sure Yule would say 'print it'. But as Burnside & Fisher had a controversy on this matter, I am inclined to think Burnside might be asked to review it—and after that, we must, you must, judge. This I suggest in the interests of truth only, since the matter is really difficult.

'Print it' had been Yule's response to the "Foundations"; see Sect. 5 above. Yule, like Baker—White and Jeffreys too—was a fellow of St. John's and the only professed statistician in Cambridge. It may be significant for the way the Burnside–Fisher relationship was remembered that Baker already saw a "controversy" although there had not been one either in what the principals wrote publicly or, it seems, in what they wrote privately—see Sect. 4 above and Sect. 8 below.

The judgement was soon made and on the 25th Baker wrote to Fisher explaining that his paper had been sent to Burnside, "not to determine whether it should be printed, but because you would probably like to feel it had been carefully read by a fresh mind." Baker enclosed some notes from Burnside and invited Fisher to discuss them with Burnside and perhaps "modify" his paper; there was no obligation to do this but "it is fair to say that Dr. Burnside says he will feel the necessity of writing a short note to yours, after it has been published, if it appears as it stands." Baker reflected that "on a matter which is really so difficult, a clash of opinion is surely, in the long run, likely to be a gain." He added that he had had considerable correspondence with Burnside in "times past" on probabilities. Fisher modified the paper only by adding a "prefatory note". On the 29th after congratulating Fisher on the birth of his daughter Burnside thanked him for the typescript of the note. On April 7th he told Fisher he was sending Baker a note on the idea of frequency.

The article as published begins with the prefatory note and is followed by Burnside's two page note, "On the idea of frequency." This takes off from two sentences from the "Foundations" (p. 311) on how the "reduction of data" is accomplished:

> This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample. The law of distribution of this hypothetical population is specified by relatively few parameters which are sufficient to describe it exhaustively in respect of all qualities under discussion.

(Virtually the same words appear in the new paper (p. 701).) These are not the most relevant sentences for they do not mention "frequency" and for Burnside (1925b, p. 726) the issue was whether the conception of frequency of a class in a finite collection can be "legitimately extended to the case of a non-finite collection." With decades of post-Cantor work in mind he wrote, "there will be probably be many who will wonder why such a question should be asked at all when the answer must so obviously be in the negative" but "there are writers in theoretical statistics who assume that the answer must be equally obviously in the affirmative."

Burnside recognises two approaches to the matter: "one may either attempt to construct a non-finite collection in which a given class has a given frequency:or starting from a given non-finite collection one may attempt to determine the frequency of particular class." He described the difficulty associated with the first course and to illustrate the difficulty associated with the second course he considered the frequency of even positive integers in the totality of all positive integers and showed how this frequency can take any value. He concludes, "Until a unique value, which cannot be disputed, has been found for this frequency is there any reason for supposing that the phrase 'the frequency of all even integers in the totality of all integers' has any meaning at all?"

Fisher's (1925b, p. 700) note begins, "It has been pointed out to me that some of the statistical ideas employed in the following investigation have never received a strictly logical definition and analysis" and goes on to mention Burnside—the only indication in Fisher's published writings of any interaction between the two men.

> It was no part of my original intention to deal with the logical bases of these ideas, but some comments which Dr. Burnside has kindly made have convinced me that it may be desirable to set out for criticism the manner in which I believe the logical foundations of these ideas may be established.

As Fisher had already seen something of Burnside's criticism, he may have thought that the exposition of the logical foundations that followed met Burnside's objections.

Fisher (1925a, p. 700) contends, "The idea of an infinite hypothetical population is, I believe, implicit in all statements involving mathematical probability." This time, instead of a die, there is an illustration from genetics:

> If, in a Mendelian experiment, we say that the probability is one-half that a mouse born of a certain mating shall be white, we must conceive of our mouse as one of an infinite population of mice which might have been produced by that mating. ... The proportion of white mice in this imaginary population appears to be the actual meaning to be assigned to our statement of probability.

Fisher believed he was making explicit the thinking behind the practice of "practical statisticians." In his own work, he (1918, p. 400n) had considered an "infinite fraternity," that is, "all the sons which a pair of parents might conceivably have produced," a formulation echoing Galton (1889, p. 125) who had imagined an "exceedingly large Fraternity, far more numerous than is physiologically possible" from which random samples are taken.

Fisher (p. 700) sets out the "manner in which I believe the logical foundations of these ideas (of frequency curve, infinite hypothetical population, random sampling, etc.) may be established."

> Imagine a population of N individuals belonging to s classes, the number in class k being $p_k N$. The population can be arranged in order in $N!$ ways. Let it so be arranged and let us call the first n individuals in each arrangement a sample of n. Neglecting the order within the sample, these samples can be classified into several possible types of sample according to the number of individuals of each class which appear. Let this be done, and denote the proportion of samples which belong to type j by q_j, the number of types being t. Consider the following proposition.
>
> Given any series of proper fractions $P_1, P_2, \ldots, P_s$, such that $S(P_k) = 1$, and any series of numbers $\eta_1, \eta_2, \ldots, \eta_t$, however small, it is possible to find a series of proper fractions $Q_1, Q_2, \ldots, Q_t$, and a series of positive numbers $\epsilon_1, \epsilon_2, \ldots, \epsilon_s$, and an integer N_0, such that if
>
> $$\begin{aligned} &N > N_0 \\ &\text{and} \quad |p_k - P_k| < \epsilon_k \quad \text{for all values of } k, \\ &\text{then will} \quad \left|q_j - Q_j\right| < \eta_j \quad \text{for all values of } j. \end{aligned}$$
>
> I imagine it possible to provide a rigorous proof of this proposition, but I do not propose to do so. If it be true, we may evidently speak without ambiguity or lack of precision of an infinite population characterised by the proper fractions P, in relation to the random sampling distribution of samples of a finite size n.
>
> It will be noticed that I provide no definition of *a* random sample, and it is not necessary to do so. What we have to deal with in all cases is a random sampling distribution of samples, and it is only as a typical member of such a distribution that a random sample is ever considered.

This remarkable passage of mock epsilonics and its combination of "I imagine it possible to provide a rigorous proof of this proposition" with complete indifference to providing one is revealing of Fisher's attitude to mathematics. It is of a piece with what Mahanalobis (1938, p. 266) reports of Fisher's attitude, "The explicit statement of a rigorous argument interested him, but only on the important condition that such explicit demonstration of rigour was needed. Mechanical drill in the technique of rigorous statement was abhorrent to him, partly for its pedantry, and partly as an inhibition to the active use of the mind." The prefatory note is the longest sustained discussion of probability in Fisher's writing before his 1956 book *Statistical Methods and Scientific Inference*.

Burnside was not the first—or the last—to query Fisher's conception. In July 1922 Fisher sent the "Foundations" to N. R. Campbell, the experimental physicist and philosopher of science who had written on probability in the *Philosophical Magazine* (1922) and previously in a book (1920). Fisher (Bennett 1990, p. 61) judged Campbell's views to be in fundamental agreement with his own: the important point was that both considered chance to be "a physical property of the material system concerned." However, Campbell (Bennett, p. 58) objected to Fisher's formulation that it had said nothing about randomness—referring to the example of the die, he pointed out that Fisher's statement is compatible with the regular appearance of a five; Campbell also complained that an infinite number of throws "means nothing" to the experimenter. Fisher (Bennett, p. 61) conceded the point about randomness but he considered the Ellis–Cournot frequency theory "sound." Fisher probably knew of these authors and quoted their words from Keynes's *Treatise on Probability* (1921, ch. VIII) which criticised their position. In his review of the *Treatise* (Fisher 1922/3, pp. 46–7) objected to what he called its "psychological" notion of probability as measuring the "degree of rational belief to which a proposition is entitled in the light of given evidence." By contrast, "To the statistician probability appears simply as the ratio which a part bears to the whole of a (usually infinite) population of possibilities." For more on Fisher and Keynes see Aldrich (2008).

Some of the Burnside issues reappeared in a (public) debate between Fisher and Harold Jeffreys whose view of probability resembled that of Keynes: see Howie (2002) and Aldrich (2005b). Jeffreys first wrote about probability and the frequency theory—in the form of the Venn limit—in Wrinch and Jeffreys (1919) but he returned to the subject in his *Scientific Inference* (1931). Jeffreys was as alarmed as Burnside by Fisher's notion of ratios taken from infinite hypothetical population, writing (p. 220n) that "with what looks like the courage of despair... Fisher (1922a) says that the ratio is perfectly definite." Later Jeffreys (1933, p. 533) remarked, "That a mathematician of Dr. Fisher's ability should commit himself to the statement that the ratio of two infinite numbers has an exact value can only be regarded as astonishing." Perhaps it is a symptom of the non-existence of a well-defined field that Jeffreys did not come across Burnside's work; there was another parallel between the two for Jeffreys (1931, pp. 66ff.) extended Gauss's theory of errors to accommodate unknown precision, like Burnside (1923b)—see Sect. 4. For his magnum opus, the *Theory of Probability* (1939), Jeffreys studied the literature more thoroughly and this (p. 304) has an approving reference to Burnside's writing on frequency.

8 The end of the correspondence April 1926

Burnside became increasingly frustrated as he waited for a response to his questions about frequency in infinite collections. On August 7th he wrote telling Fisher that "On the idea of frequency" had been published and, picking up a remark in Fisher's prefatory note, asked, "Is it not at least possible that the idea of a infinite hypothetical population has grown up in the minds of practical statisticians and has been put at the basis of recent work before it has been submitted to exhaustive criticism?" When Fisher said he had not yet seen the note Burnside wrote on that he would wait until Fisher had had a chance to look at it, though he added, “I feel almost sure that there is more between us than a difference of language.” Burnside was still waiting in November and he sent Fisher an off-print of the note. He reiterated his doubts about frequency in non-finite collections and closed with a question

> I think you will find that the answer of any one who has studied the theory of aggregates will be—No. Since the question is obviously concerned with non-finite aggregates, should one attach any importance to the answer of a person who has not studied the theory of aggregates?

The March 1926 issue of the *Philosophical Magazine* carried a note by Burnside, “On the ‘hypothetical infinite population’ of theoretical statistics.” This focusses on the same two sentences in the “Foundations” and Burnside says he is concerned with giving precision to the conception of the “hypothetical infinite population” given that nothing further in Mr. Fisher’s memoir occurs to achieve this!

One passage in Burnside (1926, p. 673)—actually one sentence—played a part in the breakdown of the relationship between Burnside and Fisher. It is the rather contorted punch-line which comes in the second sentence:

> When the “hypothetical infinite population” comes under head (ii.) [countably infinite] it is not possible to obtain a unique value, other than zero, for the frequency-ratio of a specified class by calculating the frequency ratio f_N of the class in a collection of N from the population, and determining the limit of f_N as N increases. This statement is not inconsistent with the further statement that “the frequency-ratio of a specified class in a hypothetical infinite population” is a phrase without meaning.

It is not clear when Burnside wrote the note for he had a slight stroke in December 1925 which put him out of action for a time—see Everett et al. (2004, p. 106). However by April he was writing to Baker about a paper for the *PCPS* and to Fisher about the theory of statistical estimation. The correspondence with Fisher resumes abruptly on the 21st with a query about the explanation of consistency in Fisher (1925b)—“Is it possible for a statistician to arrive at a numerical value of what you denote by T_∞ [the probability limit of the sequence of estimators T_n]?”—but on the 24th he is back to the old topic, “What I really want to know is that I am correctly representing you in what I say of a ‘hypothetical infinite population.’”

Burnside’s letter of the 27th was not business as usual. The last letter to Fisher, it is full of injury and has an ultimatum:

> Dear Dr. Fisher
>
> In your letter received this morning the postscript runs verbatim thus:—
> May I call your attention to this place in your Phil. Mag. note
> This statement is not inconsistent with the further statement.... “.....” is a phrase without meaning. Is not this a rather underhand way of suggesting, without asserting, that there something in your previous reasoning which proves or implies that the phrase in question has no meaning, whereas nothing of the sort is —?

> I have put a dash between the last word but one–is–and the question mark because I cannot be certain what the word is. With that exception I believe the passage is reproduced verbatim from your letter.
> I absolutely object to the use of the word "underhand" in connection with anything I have done; and unless it is unequivocally withdrawn this correspondence must cease.

Fisher evidently replied for Burnside quoted part of the letter in a desperate and incoherent appeal to Baker dated two days later:

> Excuse this digression. In the course of a discussion by post on a purely scientific point with a correspondent whose name I wish to withhold, he (the correspondent) wrote the following sentence. "Is not this a rather underhand way of suggesting, without asserting, that there something in your previous reasoning which proves or implies that the phrase in question has no meaning, whereas nothing of the sort is proved?"
>
> If the words "rather underhand" were omitted it is perfectly possible that there might be something in my previous reasoning which I thought proved that the phrase in question had no meaning, while my correspondent thought it proved nothing of that sort; so that the asking of the question, omitting "rather underhand" would not necessarily carry an offensive meaning. The introduction of the words "rather underhand" necessarily give the question an offensive meaning.
>
> I accordingly wrote to him to say that unless the word "underhand" was unequivocally withdrawn the correspondence must at once cease.
>
> The sentence "In suggesting it, I assume, like other men, you occasionally, in action, fall below your own standards: but I do not forget that the more intellectual the work, the higher the standards are required to be" occurs and no withdrawal is made in the letter that followed mine asking for a withdrawal. This seems to me even more offensive than the original question.
> It is a relief to put the whole matter before an independent third party but I am I don't see how you can possibly help me.

How Baker reacted is unknown but the next day Burnside returned to the mathematical business without mentioning the "digression." There were no more letters to Fisher. Everett et al. (2004, p. 104) suggest that Forsyth (1928a, p. xvii) was alluding to this incident when he recalled Burnside's detestation of controversy. I think he was, although some details do not fit, e.g. it was not "a question of priority" as Forsyth says but a question of honour, originating in an accusation of cheating. However, if Baker told Forsyth of the Fisher incident he may have interpreted the dispute as fundamentally one about priority; see Sect. 7 above. Whether the incident—and Burnside's dealings with Fisher generally—illustrates detestation of controversy is another matter. He may not have enjoyed controversy as an art-form as Pearson and Fisher did—but he liked to win.

Burnside died on 21 August 1927 aged 75 and Forsyth (1928a, p. xiv) records that "the last year of his life was marked by failing health." We have put aside some topics to follow the Fisher story to its conclusion. They are considered in the next section.

9 Probability and the *Theory of Probability* 1919–1925

While he engaged with the statisticians Burnside worked on other parts of the probability project—the book and a few satellite papers. The book was never finished but it "contained all the issues which he proposed to discuss", according to Forsyth (1928b, p. v) who prepared the manuscript for publication. Forsyth also relates, "So far as can be remembered by Mrs. Burnside, the draft was written at intervals before the middle of 1925" and that after his stroke "he occasionally longed to return to the draft, so as to make additions and amplifications: but the necessary strength was lacking."

In probability proper Burnside published nothing as provocative as the '*t* paper' and, in any case, there were no specialists to provoke—no counterparts to the biometricians or to Fisher, the statistical crusader. In England probability per se was left to the philosopher-logicians for whom it was a branch of logic and to scientists with philosophical leanings. Of these only Jeffreys might have had a useful exchange of ideas with Burnside. Cayley and Sylvester had solved probability problems set in the *Educational Times* because pure mathematicians liked puzzles, not because probability was their business. In applied mathematics there was interest and capacity, but no specialisation or exclusiveness. This was manifest in the subject of "random flights" (random walks in higher dimensions) to which Burnside's note, "The Problems of Random Flight and Conduction of Heat" (1924d), was a typical contribution. This was another efficient self-contained production with a limited objective—to connect a result on random flights from Rayleigh (1919) and a well-known piece of applied mathematics, the heat equation. It used no specialised probability technique and made no connection with any topic in the book Burnside was writing. The note seems to have gone unnoticed but that was the fate of most of the papers described in Dutka's (1985) review of the subject; the papers did not coalesce into a subject as each writer started from scratch using only standard analytical techniques. Pearson (1905) was one of these writers and Fisher could be reckoned another as he (1922c and 1930) tackled related problems involving diffusion in population genetics. The topic was eventually gathered into the new field of stochastic processes being developed by Continental mathematicians with heat equations re-appearing as "forward equations."

Burnside's *Theory of Groups of Finite Order* (1897) had been a new kind of book for England and so, in a way, was the *Probability* for no book had treated combinatorial probability, geometrical probability and the theory of errors together. In the event, a book with a similar specification appeared before the *Probability*: the *Introduction to Mathematical Probability* (1925) by the Harvard geometer Julian Coolidge. The premiss of *Finite Groups* had been that its subject "had attracted but little attention in this country" (p. vi) but that great advances were being made elsewhere; that too was Coolidge's (1925, p. v) premiss. However Burnside's *Probability* did not look around but backwards and offered a close examination of old issues. The developments discussed by von Plato (1994), Bingham (2000) and Shafer and Vovk (2006)—the foundations of modern probability—were simply out of Burnside's sight. When David (1960) reviewed the Dover reprint of the *Probability* she compared it not with contemporary productions like Borel (1925) but with Crofton's old encyclopedia article. From his references it seems that Burnside used only Poincaré (1912, original edition 1897) and Bertrand (1888). More up-to-date works were not entirely unknown in

England; Markov (1912) and other Russian contributions were trumpeted in Keynes's *Treatise*—see Aldrich (2008)—but that too was out of sight.

In design the *Probability* resembled the *Encyclopaedia Britannica* articles on probability by Crofton (ninth edition, 1885) and Edgeworth (eleventh edition, 1911). The chapters are: Introduction; Direct Calculation of Probabilities; Indirect Methods of Calculating Probabilities; Methods of Approximation; Probability of Causes; Probabilities connected with Geometrical Questions; Theory of Errors; Gauss's Law of Errors. Notes were planned "elucidating or establishing statements in the text" but only one was written. There is no author's preface and no statement of his intentions. Forsyth contributed an index and a list of "cognate" papers. He missed two probability papers as well as two on difference equations that might be considered "cognate." Only one paper seems to have rehearsed material for the book, the error paper of 1923, and it produced the only drama, the reverse described in Sects. 3 and 4 above. Otherwise Burnside moved from topic to topic in no particular order with no widening of vision or deepening of technique.

The *Probability* takes its examples from games of chance, geometry and the combination of observations though it notes that "a good deal of modern molecular physics is bound up with certain calculated probabilities." (1928, p. 13). Coolidge's much larger book has chapters on the theory of gases and on insurance as well as material on statistics—both English (Yule on correlation) and Continental (Lexis on dispersion). Coolidge founded probability on a combination of von Mises and John Stuart Mill while Burnside stuck to the classical approach and mentioned no others. He (p. 4) replaces Poincaré's formulation in which a trial has a number n of equally likely results with one in which "each two of the results are assumed to be equally likely"; the one elucidatory note Burnside finished discusses the need for this refinement. Burnside refers to Poincaré but the formulation goes back to Laplace and there are some comments on Burnside's refinement from Dale (1995, p. 136). The assumption of "equal likelihood" receives considerable attention as the whole construction depends upon it; Burnside (1924e) is a separate note on the subject.

Chapter II on the "Direct Calculation of Probabilities" presents the standard combinatorial material, though with fresh examples. Chapter III on "indirect methods" treats problems of the duration of play; some of Burnside's results were rediscovered under the heading of the "random walk in the presence of absorbing barriers"—see Kac (1945). The organising principle of the chapter is that "the required probability can often be determined by obtaining a finite difference equation which it must satisfy." (pp. 26–7). Burnside wrote two papers (1923a and 1924b) on difference equations. Chapter IV on "Methods of Approximation" treats what was elsewhere being called the "central limit theorem," the normal approximation to the binomial and the multinormal approximation to the multinomial; cases relevant to the theory of errors are treated later. The paper "On the Approximate Sum of Selected Terms from the Multinomial Expansion" (1925a) was a spin-off from this chapter. Chapter V on the "Probability of Causes," treats the "Bayes formula" and emphasises the importance of the form of prior assumed. Burnside's first probability paper "On the probable regularity of a random distribution of points" (1919) had treated a geometrical problem and chapter VI on "Probabilities connected with Geometrical Questions" treated several more. His 1924 paper "On the phrase 'equally probable'" was another study in geometric probability; based on an example from Poincaré, it had no connection with Bayesian inference.

The final two chapters on errors evoke a pre-Fisher state of innocence into which knowledge of the Student result has entered. The first chapter begins by considering what can be learnt about the "unknown quantity" α when different assumptions are made about the properties (non-stochastic) of the measurement errors. It then presents Gauss's derivation of the normal distribution as the distribution that gives the arithmetic mean as the most probable value of the unknown quantity, assuming a uniform prior. Burnside (pp. 84–5) argues that within the range of the observations it is "quite unrealistic to assume rapid variation of $p(\alpha)$" and as the posteriors do not vary much when the priors are tied down in this way we may as well choose the simplest prior, viz. the uniform. Burnside then gives a central limit theorem argument for the errors to be normally distributed. The propositions do not contradict one another and the proofs are tidy but it is hard to see any architecture in the way everything is assembled. The chapter could have been finished by early 1923 and so could most of the following chapter. This reviews the main properties of "Gauss's law of errors" and then in Sect. 36 on the "combination of observations" it asks "what deductions can be drawn from a given set of determinations?" The answer (p. 96) takes the form of the sampling distribution of (Burnside's variant of) the t-ratio with a statement of the limits between which the unknown value is as likely as not to lie; the numerical values of the limits are those given in the 1923 article but the proof is the one he had outlined to Fisher in August 1923. To all appearances Burnside simply took out the old Bayesian Sect. 36 and put in a new sampling distribution argument. The reference to Student never materialised.

Burnside's "On an integral connected with the theory of probability" (1924a) has an affinity with his proof of Student's result but, as with the partial difference papers, there is no motivation for the generalisation: the word "probability" appears only in the title. Let x_{ij} ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$; $m > n$) denote n sets of m real variables then "it is proposed to determine the integral of the differential" $\prod_{ij} dx_{ij}$ over the range given by the inequalities

$$a_{jk} \leq \prod_{i=1}^{m} x_{ij}x_{ik} \leq a_{jk} + \delta a_{jk} \quad (j, k = 1, 2, \ldots, n).$$

The book was not reviewed in the statistical journals. The British science journals *Nature* (Milne-Thomson 1928) and the *Philosophical Magazine* (Anon 1928) received it respectfully, noting the author's distinction, summarising the contents and recommending it. The *American Mathematical Monthly* had a considered review by a probability specialist, Edward Molina, remembered today for his (1930, 1931) studies of Laplace and Bayes. Molina (1929) was disappointed with the book. He (p. 95) was not impressed by the critique of Poincaré and he noted "inconsistencies" in the treatment of the theory of errors both within the book and between the book and the author's articles. He points out that a Bayesian analysis for the case of two parameters requires "an a priori function involving both of these unknowns" and that "Student's formula" is the solution to a "totally different problem" and will not give the posterior distribution "*unless arbitrary assumptions are made*" such as a uniform prior for the unknown true value and a prior for the precision h proportional to $1/h$. So much so clear, but Molina seemed to think that the two Burnside 1923 pieces and the Fisher piece were *all* concerned with the "a posteriori problem." (Forsyth's list of "cognate papers" had included a reference to Fisher (1923)). Molina goes on to state:

> R. A. Fisher is, in the opinion of the reviewer far from giving a legitimate treatment of the problem. However, we are indebted to Burnside for a short rigorous derivation of the Student formula.

Perhaps if Forsyth had noticed Burnside's note on Bayes, Molina might have found some interesting things to say about it. Molina refers to Pearson's "fundamental problem of practical statistics" in his 1931 article on Bayes but only for the eloquence of the phrase!

From the perspective of what we have of the *Theory of Probability* Burnside's transactions with the statisticians were a great waste of time. When Burnside first met the modern theory of statistics in the summer of 1923 he went at it with such energy that he may have intended to treat it in his book. What to say must have been a problem and it seems unlikely that he ever envisaged treating the chimeras of '24 and '25, Pearson's solution to the fundamental problem and Fisher's hypothetical infinite population. A reference would have sufficed for Student but, as summer 1923 turned into winter, Burnside seems to have persuaded himself that no acknowledgment was due; see Sect. 4 above.

In the *Probability* Burnside produced a concise and sometimes acute re-presentation of nineteenth century probability. There were some novelties but they did not stimulate new work; the book had readers but they seemed only to show themselves to point out that some new result such as that of Irwin–Sect. 4 above—or of Kac was already in Burnside!

10 Concluding remarks

The *Collected Papers* project stands as a monument to Burnside's work in group theory but his work in probability and statistics has seemed too slight to justify much interest in him, or in the circumstances in which he wrote. I have tried to show that these minor writings are more interesting when put together and read with his letters to the man who became the dominant figure in statistics in the first half of the twentieth century, Ronald Fisher.

Neumann et al. (2004, p. 32) describes Burnside as working in isolation from other group theorists, "possibly even more so than was normal for his times." In probability Burnside was even more isolated, retired to the country with a few books and articles though with skill, energy and time in abundance. From the Continent it may well have seemed that the whole of England had retired to the country long ago—it contributed so little to probability. By 1920, however, decades of separate development had produced an English subject, biometry/statistics, and an establishment to serve it. When Fisher arrived with the key to the new and inaccessible subject of "Statistics" Burnside's isolation was broken.

Fisher too was isolated, exiled even: he was separated from the biometricians and while others recognised his brains they did not recognise his ideas. The relationship between Fisher and Burnside started with promise but produced mainly disappointment for their minds failed to meet. Burnside began by wanting to get to the bottom of statistics and finished wanting an acknowledgment that *he* was right about frequencies in infinite populations; he achieved neither. For Fisher, Burnside was a catch; anybody interested in his work would have been but Burnside was not just anybody. However the only positive outcome of the statistical education of William Burnside was one note criticising Pearson; otherwise there was a lot of pestering over a matter Fisher did not want to discuss and a book with a lot of unreconstructed thoughts about statistical inference. Fisher's larger overtures to Cambridge were no more successful—the chair in mathematical statistics came in 1961. Fisher did well enough eventually: election to the Royal Society in 1929, advancement to Pearson's chair in 1933 and return to Cambridge as professor of genetics in 1943; the Statistical Society was remodelled to accommodate his version of statistics—see Aldrich (2008)—and the international Biometric Society made him its first President.

Burnside's engagement with the statisticians was a curious business: Burnside found errors in the reasoning of Pearson and Fisher on what they considered fundamental points, yet his discoveries seem to have had no consequences and his papers never became part of "the literature." In the case of Pearson's solution to the "fundamental problem of practical statistics" this is not surprising for it was already clear when Burnside intervened that this was *not* going to be "of far reaching importance in the theory of probabilities." Pearson's paper was quickly forgotten although it has come back into circulation as an attempt at Bayes exegesis and with it has come Burnside's note. Fisher's "Foundations" *did* produce a literature and its ideas on infinite hypothetical populations were practically immortalised by being embedded in the introductory chapter of *Statistical Methods for Research Workers* (1925a), the 14th edition of which Fisher was preparing when he died in 1962. What is less clear is how much Fisher's ideas on probability mattered to his work in statistical theory and how seriously they were taken by other statisticians. Fisher (1925a, p. 5) may have believed and continued to believe that a "frequency distribution" specifies the fractions of the infinite population assigned to the different classes but there is no material difference between his statement of the normal distribution later in the book (1925a, p. 46) and Burnside's (1923a, p. 482) statement of "Gauss' law" (above Sect. 3)). It is hard to know what other statisticians thought because the matter was so little discussed but I suspect Fisher's more sophisticated readers treated the "hypothetical infinite population" as an idiosyncrasy of expression comparable to Fisher's use of S instead of Σ for summation—see Aldrich (2003)—and for an analysis of probability went to von Mises and others. Burnside's notes appear to have been noticed only by Jeffreys and the *Theory of Probability* was hardly a mainstream work.

There are two footnotes to this story: a connection of a kind between Burnside and Fisher and the circumstances of the re-issue of the *Probability* in 1959. The connection is that Fisher's theory of experimental design generated problems in group theory. Fisher first worked on experimental design in his Burnside years—the subject is announced in his (1925, Sect. 48). Later he (1942) brought groups into the theory; he referred to no books and the only book on groups in his personal library was Hilton

(1908) essentially an introduction to Burnside's *Finite Groups*. Burnside had reviewed the book and his concluding words were, "The appearance of this book ought to do a good deal to reduce the apathy with which its subject is treated in this country." (Burnside 1908, p. 336). In 1955 Dover Books reprinted Burnside's book on groups and in 1959 it re-issued the *Probability*. The latter was not a classic but it was by Burnside and probability had recently come to life. For this Feller's *Introduction to the Theory of Probability and its Applications* (1950) was largely responsible but Feller's book came out of an intellectual movement that Burnside—or even Coolidge which was also re-issued by Dover—had no part in.

References

Burnside's letters to Fisher are in the Barr–Smith Library of the University of Adelaide, as are the two letters from Baker about the publication of Fisher (1925). The other Baker letters are in St. John's College Cambridge. Burnside's articles, except for "On Bayes' Formula" (1924c), are reproduced in Neumann, Mann and Tompson (eds.) (2004).

Aldrich, J. (1997) R. A. Fisher and the Making of Maximum Likelihood 1912–22, *Statistical Science*, **12**, 162–176.

Aldrich, J. (2003) The Language of the English Biometric School, *International Statistical Review, 71,* 109–130.

Aldrich, J. (2005a) Fisher and Regression, *Statistical Science*, **20**, 401–417.

Aldrich, J. (2005b) The Statistical Education of Harold Jeffreys, *International Statistical Review*, **73**, 289–308.

Aldrich, J. (2007) The Enigma of Karl Pearson and Bayesian Inference, paper based on a talk given at the *Karl Pearson Sesquicentenary Conference*, the Royal Statistical Society March 2007.

Aldrich, J. (2008) Keynes among the Statisticians, *History of Political Economy*, **40**, 265–316

Anon. (1928) Notices respecting New Books: *Theory of Probability*, by the late William Burnside, *Philosophical Magazine*, **6**, 943.

Barrow-Green, J. (2004) Burnside's Applied Mathematics, pp. 63–70 of Neumann, Mann and Tompson (2004).

Bayes, T. (1763) An Essay towards Solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society,* **53**, 370–418.

Bennett, J. H. (1983) (ed) *Natural Selection, Heredity, and Eugenics: including Selected Correspondence of R.A. Fisher with Leonard Darwin and Others*, Oxford, University Press.

Bennett, J. H. (1990) (ed) *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*, Oxford, University Press.

Bertrand, J. (1888) *Calcul des Probabilités*, Paris: Gauthier-Villars.

Bingham, N. H. (2000) Measure into probability: From Lebesgue to Kolmogorov, *Biometrika* , **87**, 145–156.

Borel, E. (1925) *Traité du Calcul des Probabilités et de ses Applications*, Paris: Gauthier-Villars

Bowley, A. L. (1935) Discussion of Fisher's Logic of Inductive Inference, *Journal of the Royal Statistical Society,* **98**, 55–57.

Box, J. F. (1978) *R. A. Fisher: The Life of a Scientist*, New York: Wiley.

Broemeling, L. & A. Broemeling (2003) Studies in the History of Probability and Statistics XLVIII: The Bayesian Contribution of Ernest Lhoste, *Biometrika*, **90**, 720–723.

Bru, B. (1996) La Probléme de l'Efficacité dur Tir à l' École de Metz, *Mathématiques, Informatiques et Sciences Humaines*, **34**, 25–38.

Burnside, W. (1888) On the Partition of Energy between the Translatory and Rotational Motions of a Set of Non-homogeneous Elastic Spheres, *Transactions of the Royal Society of Edinburgh*, **33**, 501–507.

Burnside, W. (1889) On a Simplified Proof of Maxwell's Theorem, *Proceedings of the Royal Society of Edinburgh*, **15**, 106–108.

Burnside, W. (1897) *The Theory of Groups of Finite Order*, second edition in 1911, Cambridge: Cambridge University Press. (Second edition reprinted in 1955 by Dover, New York.)

Burnside, W. (1908) Review of *An Introduction to Groups of Finite Order*, by H. Hilton, *Mathematical Gazette*, **4**, 335–336.
Burnside, W. (1919) On the Probable Regularity of a Random Distribution of Points, *Messenger of Mathematics*, **48**, 47–48.
Burnside, W. (1923a) On Errors of Observation, *Proceedings of the Cambridge Philosophical Society*, **21**, 482–487.
Burnside, W. (1923b) The Solution of a Certain Partial Difference Equation, *Proceedings of the Cambridge Philosophical Society*, **21**, 488–491.
Burnside, W. (1923c) On Errors of Observation, *Proceedings of the Cambridge Philosophical Society*, **22**, 26–27.
Burnside, W. (1924a) On an Integral Connected with the Theory of Probability, *Messenger of Mathematics*, **53**, 142–144.
Burnside, W. (1924b) On a Partial Linear Difference Equation, *Messenger of Mathematics*, **53**, 161–165.
Burnside, W. (1924c) On Bayes' Formula, *Biometrika,* **16**, 189.
Burnside, W. (1924d) The Problems of Random Flight and Conduction of Heat, *Proceedings of the Cambridge Philosophical Society*, **22**, 168–167.
Burnside, W. (1924e) On the Phrase "equally probable" *Proceedings of the Cambridge Philosophical Society*, **22**, 669–671.
Burnside, W. (1925a) On the Approximate Sum of Selected Terms from the Multinomial Expansion, *Messenger of Mathematics*, **54**, 189–192.
Burnside, W. (1925b) On the Idea of Frequency, *Proceedings of the Cambridge Philosophical Society*, **22**, 726–727.
Burnside, W. (1926) On the 'Hypothetical Infinite Population' of Theoretical Statistics, *Philosophical Magazine Series 7*, **1**, 670–674.
Burnside, W. (1928) *Theory of Probability*, Cambridge: Cambridge University Press. (Reprinted in 1959 by Dover, New York.)
Campbell, N. R. (1920) *Physics: The Elements*, Cambridge, Cambridge University Press.
Campbell, N. R. (1922) The Measurement of Chance, *Philosophical Magazine*, **44**, 67–79.
Coolidge, J. L. (1925) *An Introduction to Mathematical Probability*, Oxford: Clarendon Press. (Reprinted in 1962 by Dover, New York.)
Crofton, M. W. (1885) Probability, *Encyclopaedia Britannica,* ninth edition, 768–788. London.
Dale, A. I. (1991) *A History of Inverse Probability from Thomas Bayes to Karl Pearson*, expanded second edition, 1999. New York: Springer.
Dale, A. I. (tr. & ed.) (1995) *Pierre-Simon Laplace*: *Philosophical Essay on Probabilities*, New York: Springer.
David, F. N. (1960) Review of Burnside's *Theory of Probability, Journal of the Royal Statistical Society. Series A*, **123**, 339.
Dutka, J. (1985) On the Problem of Random Flights, *Archive for History of Exact Sciences*, **32**, 351–375.
Edgeworth, F. Y. (1911) Probability, *Encyclopaedia Britannica,* eleventh edition, vol. 22, 376–403. London.
Edgeworth, F. Y. (1921) Molecular Statistics, *Journal of the Royal Statistical Society*, **84**, 71–89.
Edwards, A. W. F. (1978) Commentary on the Arguments of Thomas Bayes, *Scandinavian Journal of Statistics*, **5**, 116–118.
Edwards, A. W. F. (1997) What did Fisher Mean by "Inverse Probability" in 1912–1922? *Statistical Science*, **12**, 177–84.
Eisenhart, C. P. (1979). On the Transition from 'Student's z' to 'Student's t', *American Statistician*, **33**, 6–10.
Everett, M. G., A. J. S. Mann and K. Young (2004) Notes on Burnside's Life, pp. 89–109 of Neumann, Mann and Tompson (2004).
Feller, W. (1950) *An Introduction to the Theory of Probability and its Applications, vol. 1*, New York: Wiley.
Fisher, R. A. (1915) *The Mathematical Theory of Probabilities and its Application to Frequency Curves and Statistical Methods*, second edition 1922, New York: Macmillan.
Fisher, R. A. (1915) Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population, *Biometrika*, **10**, 507–521.
Fisher, R. A. (1918) The Correlation between Relatives on the Supposition of Mendelian Inheritance, *Transactions of the Royal Society of Edinburgh*, **52**, 399–433.
Fisher, R. A. (1921) On the 'Probable Error' of a Coefficient of Correlation Deduced from a Small Sample, *Metron*, **1**, 3–32.

Fisher, R. A. (1922a) On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society, A*, **222**, 309–368.
Fisher, R. A. (1922b) The Goodness of Fit of Regression Formulae, and the Distribution of Regression Coefficients, *Journal of the Royal Statistical Society*, **85**, 597–612.
Fisher, R. A. (1922c) On the Dominance Ratio, *Proceedings of the Royal Society of Edinburgh*, **42**, 321–341.
Fisher, R. A. (1922/3) Review of J. M. Keynes's Treatise on Probability, *Eugenics Review*, **14**, 46–50.
Fisher, R. A. (1923) Note on Dr. Burnside's Recent Paper on Errors of Observation, *Proceedings of the Cambridge Philosophical Society*, **21**, 655–658.
Fisher, R. A. (1924/8) On a Distribution Yielding the Error Functions of Several Well Known Statistics, *Proceedings of the International Congress of Mathematics, Toronto,* **2**, 805–813.
Fisher, R. A. (1925a) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.
Fisher, R. A. (1925b) Theory of Statistical Estimation, *Proceedings of the Cambridge Philosophical Society*, **22**, 700–725.
Fisher, R. A. (1925c) Applications of 'Student's' Distribution, *Metron*, **5**, 90–104.
Fisher, R. A. (1930) The Distribution of Gene Ratios for Rare Mutations, *Proceedings of the Royal Society of Edinburgh*, **50**, 205–220.
Fisher, R. A. (1941) The Asymptotic Approach to Behrens's Integral, with Further Tables for the *d* Test of Significance, *Annals of Eugenics*, **11**, 141–172.
Fisher, R. A. (1956) *Statistical Methods and Scientific Inference*, Edinburgh: Oliver & Boyd.
Forsyth, A. R. (1928a) William Burnside, *Proceedings of the Royal Society A*, **117**, xi-xxv. (Reprinted in Burnside (1928) and in Neumann, Mann and Tompson (2004).). My page references are to the printing in Burnside (1928).
Forsyth, A. R. (1928b) Preface to Burnside (1928).
Galton, F. (1889) *Natural Inheritance*, London: Macmillan.
Greenwood, M. (1933) Ethel May Newbold, *Journal of the Royal Statistical Society*, **96**, 354–357.
Hald, A. (1998) *A History of Mathematical Statistics from 1750 to 1930*, New York: Wiley, 1998.
Helmert, F. R. (1876) Die Genauigkeit der Formel von Peters zur Berechnung des wahrscheinlichen Fehlers directer Beobachtungen gleicher Genauigkeit, *Astronomische Nachrichten*, **88**, 192–218 An extract is translated and annotated in H. A. David & A. W. F. Edwards (eds.) *Annotated Readings in the History of Statistics* New York: Springer (2001).
Hilton, H. (1908) *Introduction to the Theory of Groups of Finite Order*.
Hotelling, H. (1930) The Consistency and Ultimate Distribution of Optimum Statistics, *Transactions of the American Mathematical Society* , **32**, 847–859.
Howie, D. (2002) *Interpreting Probability: Controversies and Developments in the Early Twentieth Century*, New York, Cambridge University Press.
Irwin, J. O. (1931) Mathematical Theorems Involved in the Analysis of Variance, *Journal of the Royal Statistical Society*, **94**, 284–300.
Irwin, J. O. (1934) On the Independence of the Constituent Items in the Analysis of Variance, *Supplement to the Journal of the Royal Statistical Society*, **1**, 236–251.
Irwin, J. O. (1967) William Allen Whitworth and a Hundred Years of Probability (with discussion), *Journal of the Royal Statistical Society, A*, **130**, 147–176.
Jeffreys, H. (1931). *Scientific Inference*, Cambridge: Cambridge University Press.
Jeffreys, H. (1933) Probability, Statistics and the Theory of Errors, *Proceedings of the Royal Society, A*, **140**, 523–535.
Jeffreys, H. (1939) *Theory of Probability*, Oxford: University Press.
Kac, M. (1945) Random Walk in the Presence of Absorbing Barriers, *Annals of Mathematical Statistics*, **16**, 62–67.
Keynes, J. M. (1921) *A Treatise on Probability*, London, Macmillan.
Mahanalobis, P. C. (1938) Professor Ronald Aylmer Fisher, *Sankhya*, **4**, 265–272. Reprinted in W. A. Shewhart (ed) (1950) *Contributions to Mathematical Statistics*, New York: Wiley.
Markov, A. A. (1912) *Wahrscheinlichkeitsrechnung*, (translated from the Russian by H. Liebmann).
Milne-Thomson, L.M. (M.-T., L. M.) (1928) Our Bookshelf: *Theory of Probability*, by the late William Burnside, *Nature*, **129**, (August 24), 297.
McMullen, L. (1970) *Letters from W. S. Gosset to R. A. Fisher 1915–1936: Summaries by R. A. Fisher with a Foreword by L. McMullen*, printed by Arthur Guinness for private circulation.
Molina, E. C. (1929) Theory of Probability. William Burnside; A. R. Forsyth, *American Mathematical Monthly*, **36**, 94–96.

Molina, E. C. (1930) The Theory of Probability: Some Comments on Laplace's Théorie Analytique, *Bulletin of the American Mathematical Society*, **36**, 369–392.
Molina, E. C. (1931) Bayes' Theorem: An Elementary Exposition, *Annals of Mathematical Statistics*, **2**, 23–37.
Neumann, P. M. (2004) The Context of Burnside's Contribution to Group Theory, pp. 15–37 of Neumann, Mann & Tompson (2004).
Neumann, P. M., A. J. S. Mann and J. C. Tompson (eds.) (2004) *The Collected Papers of William Burnside,* 2 vols., Oxford, Oxford University Press.
Newbold, E. (1923) Note on Dr Burnside's Paper on Errors of Observation, *Biometrika*, **15**, 401–406.
Neyman, J., Pearson, E. S. (1928) On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Part I, *Biometrika*, **20A**, 175–240.
Norton, B. & E. S. Pearson (1976) A Note on the Background to and Refereeing of R. A. Fisher's 1918 Paper 'The Correlation between Relatives on the Supposition of Mendelian Inheritance', *Notes & Records of the Royal Society of London*, **31**, 151–62.
Pearson, E. S. (1936/8) Karl Pearson: An Appreciation of Some Aspects of his Life and Work, In Two Parts, *Biometrika*, **28** , 193–257, **29**, 161–247.
Pearson, E. S. (1968) Some Early Correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with Notes and Comments, *Biometrika*, **55**, 445–457.
Pearson, E. S. (1990) *'Student', A Statistical Biography of William Sealy Gosset*, Edited and Augmented by R. L. Plackett with the Assistance of G. A. Barnard, Oxford, University Press.
Pearson, K. (1892) *The Grammar of Science,* London, Walter Scott.
Pearson, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space, *Philosophical Magazine*, **2**, 559–572.
Pearson, K. (1905) The Problem of the Random Walk, *Nature*, **62**, 294.
Pearson, K. (1907) On the Influence of Past Experience on Future Expectation, *Philosophical Magazine*, **13**, 365–378.
Pearson, K. (1914) *Tables for Statisticians and Biometricians*, Cambridge: Cambridge University Press.
Pearson, K. (1920) The Fundamental Problem of Practical Statistics, *Biometrika,* **13**, 1–16.
Pearson, K. (1921) Note on the "Fundamental Problem of Practical Statistics," *Biometrika,* **13**, 300–301.
Pearson, K. (1924) Note on Bayes' Theorem, *Biometrika,* **16**, 190–193.
Pearson, K. (1931) (Editorial) Historical Note on the Distribution of the Standard Deviations of Samples of any Size Drawn from an Infinitely Large Normal Parent Population, *Biometrika*, **23**, 416–418.
Pearson, K. (1936a) Old Tripos Days at Cambridge, as Seen from Another Viewpoint, *Mathematical Gazette*, **20**, 27–36.
Pearson, K. (1936b) Method of Moments and Method of Maximum Likelihood, *Biometrika*, **28**, 34–69.
Pfanzagl, J. & O. Sheynin (1996) Studies in the History of Probability and Statistics XLIV A Forerunner of the t-Distribution, *Biometrika*, **83**, 891–898.
Plato, J. von (1994) *Creating Modern Probability*, Cambridge: Cambridge University Press.
Poincaré, J. H. (1912) *Calcul des Probabilités*, 1st edition 1896, Paris : Gauthier-Villars.
Rayleigh, J. W. (1919) On the Problem of Random Vibrations, and of Random Flights in One, Two, or Three Dimensions, *Philosophical Magazine*, **37**, 321–347.
Shafer, G. & V. Vovk (2006) The Sources of Kolmogorov's *Grundbegriffe*, *Statistical Science*, **21**, 70–98.
Soper, H. E. (1922) *Frequency Arrays: Illustrating the Use of Logical Symbols in the Study of Statistical and other Distributions*, Cambridge: Cambridge University Press.
Soper, H. E., A. W. Young; B. M. Cave; A. Lee; K. Pearson (1917) On the Distribution of the Correlation Coefficient in Small Samples. Appendix II to the Papers of Student and R. A. Fisher, *Biometrika*, **11**, 328–413.
Stewart R. M. (1920) The Adjustment of Observations, *Philosophical Magazine*, **40,** 217–227.
Stigler, S. M. (1982) Thomas Bayes's Bayesian Inference, *Journal of the Royal Statistical Society, A*, **145**, 250–258.
Stigler, S. M. (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA, and London: Belknap Press of Harvard University Press.
Stigler, S. M. (2005) Fisher in 1921, *Statistical Science*, **20**, 32–49.
Student (1908) The Probable Error of a Mean, *Biometrika*, **6**, 1–25.
Thomson, W. & P. G. Tait (1867) *Treatise on Natural Philosophy volume 1*, 2nd edition 1883, Cambridge: University Press.
Todhunter, I. (1858) *Algebra for the Use of Colleges and Schools*, Cambridge: University Press.

Todhunter, I. (1865) *A History of the Mathematical Theory of Probability: from the Time of Pascal to that of Laplace*, London: Macmillan.
Venn, J. (1888). *The Logic of Chance*, 3rd ed. (1st edition 1866), London: Macmillan.
Whitworth, W. A. (1901) *Choice and Chance*, 5th edition (1st edition 1867), Cambridge: Deighton Bell.
Wrinch, D. & H. Jeffreys (1919) On Some Aspects of the Theory of Probability, *Philosophical Magazine*, **38**, 715–731.
Yates, F. & K. Mather (1963) Ronald Aylmer Fisher 1890–1962, *Biographical Memoirs of Fellows of the Royal Society*, **9**, 91–120.
Yule, G. U. (1911) *Introduction to the Theory of Statistics* , 6th edition in 1922, London: Griffin.