UNIVERSITY OF
Southampton
School of Social Sciences

School of Social Sciences
Economics Division
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Professor A L Bowley's Theory of the Representative Method

John Aldrich

No. 0801

This paper is available on our website
http://www.socsci.soton.ac.uk/economics/Research/Discussion_Papers

ISSN 0966-4246

Key names: Arthur L. Bowley, F. Y. Edgeworth, , R. A. Fisher, Adolph Jensen, J. M. Keynes, Jerzy Neyman, Karl Pearson, G. U. Yule.
Keywords: History of Statistics, Sampling theory, Bayesian inference.

Professor A. L. Bowley's Theory of the Representative Method *

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

Abstract
Arthur. L. Bowley (1869-1957) first advocated the use of surveys–the "representative method"–in 1906 and started to conduct surveys of economic and social conditions in 1912. Bowley's 1926 memorandum for the International Statistical Institute on the "Measurement of the precision attained in sampling" was the first large-scale theoretical treatment of sample surveys as he conducted them. This paper examines Bowley's arguments in the context of the statistical inference theory of the time. The great influence on Bowley's conception of statistical inference was F. Y. Edgeworth but by 1926 R. A. Fisher was on the scene and was attacking Bayesian methods and promoting a replacement of his own. Bowley defended his Bayesian method against Fisher and against Jerzy Neyman when the latter put forward his concept of a confidence interval and applied it to the representative method.

* Based on a talk given at the *Sample Surveys and Bayesian Statistics* Conference, Southampton, August 2008. I am grateful to Andrew Dale for his comments on an earlier draft.

December 2008

1 Introduction

"I think that if practical statistics has acquired something valuable in the representative method, this is primarily due to Professor A. L. Bowley, who not only was one of the first to apply the method in practice but also wrote a very fundamental memoir giving the theory of the method." This tribute was from Neyman (1934, p. 558) but he (p. 562) also had a development to report: "since Bowley's book was written, an approach to problems of this type has been suggested ... which removes the difficulties involved in the lack of knowledge of the *a priori* probability law." Neyman did more than demonstrate an alternative to Bowley's Bayesian argument, he rejected one of the representative methods Bowley presented and, while he accepted the other, he ignored the distribution theory that Bowley had developed for it. Not much of the fundamental memoir remained.

Thirty years ago Chang (1976) considered the historical relationship between theories of sampling and general statistical theories. The present essay adopts the same perspective but focusses on Bowley and draws on recent historical research. Bowley's memoir, "Measurement of the precision attained in sampling" (1926b), was written for the International Statistical Institute and it can be seen as a part of the ISI's long conversation on the representative method; see e.g. Kruskal and Mosteller (1980). However Bowley started work on sampling long before he became involved in the ISI discussions. The date of publication is also somewhat out of time for, while the memoir appeared in the middle of the 'Fisher decade,' it looked back to Edgeworth for its ideas on inference and distribution theory. Bowley knew Fisher and knew that Fisher rejected the Bayesian approach but he was as sceptical of Fisher's alternative methods as he would be of Neyman's confidence intervals as a way around the difficulties in the Bayesian approach

Section 2 below introduces Bowley and take his early career up to the publication of the *Elements of Statistics* in 1901. Bowley first proposed the use of sampling in 1906 and for many years he saw no need for new theory; this first phase is treated in Section 3. The 1926 memoir came from Bowley's decade of high theory which began with the transformation of the *Elements* into a treatise on mathematical statistics. Bowley's conception of mathematical statistics and his first attempts to theorise about sampling are considered in Section 4. Section 5 looks at Bowley and the contemporary debate on inverse probability, the Bayesian argument. Section 6 examines the memoir with its two versions of the representative method, the theory of random selection and the theory of purposive selection. Random selection was the technique Bowley had always advocated but purposive selection was a new departure. Section 7 digresses to note a parallel Bayesian venture, Karl Pearson on sampling from a finite population. Section 8 returns to Bowley and his reaction to Neyman (1934). Section 9 concludes with some general remarks on Bowley's theory and its place in the development of sampling theory in Britain. I do not consider how Bowley used sampling in his economic and sociological investigations or what he discovered; for these aspects, see the essays in Bulmer et al. (1991).

2 The elements 1895-1901

Arthur Lyon Bowley (1869-1957) was a numerical economist and social investigator in the tradition of the London Statistical Society and its successor, the Royal Statistical Society. The general histories of Porter (1986) and Stigler (1986) and the histories of Bayesian analysis by Dale (1999), Fienberg (2006) and Aldrich (2007) describe the ideas Bowley played with and the individuals who influenced him. There are useful general accounts of Bowley's life and work by Allen & George (1957), Allen (1968), Maunder (1972), Darnell (1981) and Dale (2001) as well as a personal memoir by his daughter Agatha H. Bowley (1972). A book-length biography by Kotz and Dale (2009) is due out soon.

"Statistics are the straw out of which, I, like every other economist, have to make the bricks." These words are quoted in the *Elements* (1901, p. 8). They came from Alfred Marshall who had guided the young Cambridge wrangler away from mathematics to economics; see Groenewegen (1995, passim) for their relationship. However, Bowley's true masters were F. Y. Edgeworth (1845-1926) and, more symbolically, W. S. Jevons (1835-1882). Bowley (1934a, p. 113) gave this estimate of the two, "For actual measurement [Edgeworth] would give place to Jevons, for the theory of measurement to no economist." The investigations of foreign trade, wages, poverty, national income etc. listed in the three pages of "select" bibliography in Allen & George (1957) underline how Bowley's main business was actual measurement. However, there were two areas where he thought the theory of measurement needed attention, index numbers and sampling, and he looked to Edgeworth for inspiration, support and methods.

The Oxford Professor of Political Economy entered Bowley's life when he was appointed to lecture on Statistics at the new London School of Economics in 1895. Bowley (1934a, p. 119) recalled the circumstances:

> on Marshall's introduction I wrote to [Edgeworth] for advice on the nature and literature of the subject, and he recommended principally Venn's *Logic of Chance*, Todhunter's *History of Probability* and Lexis' *Zur Theorie der Massenerscheinungen*, to which I naturally added his 1885 paper at the jubilee meeting and his reports on Index-Numbers. From that time till his death I constantly learned from him, worked with him, and met him frequently in London and Oxford.

These titles from philosophy, mathematics and statistics indicate Edgeworth's range and nobody was better read. The jubilee paper, "Methods of statistics" (1885), with its numerous examples of significance tests was the most applicable of Edgeworth's essays, while the reports on index numbers contained the most sophisticated applications of probabilistic reasoning to economics yet seen; see Stigler (1978) and Porter (1986, pp. 253-69) for reviews Edgeworth's career in statistics and Aldrich (1992) for a discussion of his index number work. For Bowley (1934a, p. 116) Edgeworth was "the philosopher of statistics rather than the practicioner" and he made it his task was to translate the philosophy into practice. Bowley also admired Edgeworth's mathematical economics and tried to make it accessible to ordinary economists; see Darnell (1981).

Bowley the theorist of measurement was also influenced by Karl Pearson (1857-1936) and G. Udny Yule (1871-1951). Yule and Bowley had parallel lives in the Statistical Society: Bowley first contributed to the *Journal* in 1895 and Yule in 1896 and both went on to the highest offices. They represented two types of mathematical statistician–Bowley the economist with technical facility and Yule the technician who could appear anywhere and who began by tacking the statisticians' topic of the aged poor using Pearson's biometric methods. With Edgeworth, Bowley and Yule were the mathematical group in the Society; see Aldrich (2008) for more on the mathematicals. In 1895 Pearson was already established though only three years before *he* had been getting reading from Edgeworth; for this and Edgeworth's other contributions to the "English breakthrough" see Stigler (1986, ch. 9 & 10). Pearson never joined the Society but Yule was only the first of the

Pearson-trained mathematicals to settle there. There was little movement of people or ideas the other way: a *JSTOR* search produced one reference to Bowley in Pearson's writings–for a textbook treatment of Edgeworth expansions.

In his first publications Bowley concentrated on getting numbers on trade and wages and using them–on actual measurement. His only theory piece was a 1897 paper on the "Relations between the accuracy of an average and that of its constituent parts" where the average is an index number and the constituent parts are the items and their weights; the section on the "probable errors of averages" extends results in Todhunter and Edgeworth. Bowley continued to investigate the stochastic side of index numbers in his (1911a), (1926c) and (1928b). The approach was not favoured by another wrangler turned economist, J. M. Keynes (1883-1946), and he criticised it in his Adam Smith Prize essay (1909) and again in his *Treatise on Money* (1930).

In 1901 Bowley published *Elements of Statistics,* a textbook based on his lectures at the School. The 200 pages of Part I cover data (economic and demographic in the main), descriptive statistics and diagrams. There is no probability although the chapters on rounding, index numbers and interpolation contain some mathematics. The 70 pages of Part II, "Applications of the Theory of Probability to Statistics." go from the basic ideas of probability to the law of error and bivariate correlation, giving the researches of Edgeworth and Pearson their first textbook airing. “The greater part of the mathematics employed is gleaned from [Edgeworth’s] essays and the earlier authorities to whom he makes reference” wrote Bowley (1901, p. 262) although his own reading of “earlier authorities” went no further than Todhunter. The law of error covered the normal approximation to the binomial, illustrations of the central limit theorem after Edgeworth (1892) and significance tests on the difference between means after Edgeworth (1885). The treatment of correlation followed Pearson (1896).

Bowley (1901, p. 262) begins Part II by making a case for the “tedious arguments and complicated formulae” of probability. It is for the instruction of “the so-called practical men of the present day and perhaps .. some political economists”–most of the Society, in other words. Bowley’s point is that those arguments and formulae enable us to “solve problems and investigate causal relations, which, though apparently simple, must baffle direct attempts to obtain an easy solution”. The great device is the significance test and Bowley (p. 315) reflected on what a test for the existence of a change can do:

> Without some machinery of calculation of this kind we are unable to get beyond vague and general impressions of the existence of a change, but if we take care that the conditions of the calculations are satisfied, we can by the method now developed make a definite statement quite independent of personal bias, such as “either an event has hap-

> pened so improbable as to be outside the range of human experience, or the decrease shown in the series of figures in question is due to some significant change in the system of causes which produce them."

Later Bowley looked to inverse probability–the Bayesian argument–for even more "definite" statements; see below Sections 4-6.

For Bowley asymptotic normality made tests possible, a point he was still pressing ten years later when he reviewed Yule's *Introduction to the Theory of Statistics*. Bowley recognised that this went further than the *Elements*, particularly on Yule's own topics of correlation and association, but he (1911, p. 264) regretted that the "test of significance of variations" was left "so vague":

> [Yule] is contented to show by illustration that "a range of six times the standard deviation usually includes 99 per cent. or more" of the observations in ordinary cases, and 99.7 per cent. in the normal curve; he does not use Professor Edgeworth's theorem that the frequency curves of averages are normal, and therefore cannot assign definitely the improbability of a deviation. Thus the essential test for causation is left on a purely empirical basis.

Even the first edition of the *Elements* had more emphasis on distribution but the gap became a gulf when Bowley brought out a greatly expanded fourth edition in 1920; see Section 4 below.

Unlike Edgeworth, Bowley did not involve himself with the philosophy of probability and it is difficult to get any definite sense of what probability meant for him. The account of chance in the *Elements* (1901, p. 266) follows Venn's frequency theory but the probability calculations (p. 269) involve counting equally likely cases; in the 1920 revision Bowley removed all discussion of the meaning of probability. In 1901 Bowley did not discuss inverse theory or indeed inference in general, but the two cases of estimation he treats in detail have a common theme. Bowley (1901, p. 283) adapts a derivation from Merriman's *Method of Least Squares* to show how "to find, from the observations, the centre and modulus of the [error curve] from which they would come with least improbability." Bowley (p. 318ff) also reports that Pearson's (1896) formula for correlation is based on the same argument from least improbability. There was nothing unusual in the phenomenon of maximum likelihood before maximum likelihood–other instances are noted by Aldrich (1997) and Hald (1999). However, by the time Fisher was promoting maximum likelihood over inverse theory Bowley had taken to emphasising inverse theory; see Sections 4-6 below.

3 The method of sampling 1901-15

The "method of sampling" is "persistently neglected, and even when it is used the test of precision is ignored." So Bowley (1906, p. 553), as president of the Economic

Science and Statistics Section of the British Association for the Advancement of Science, informed his audience. His larger theme was the importance of probabilistic reasoning echoing the remarks in Part II of the *Elements*. On this occasion, after sketching the history of such reasoning, Bowley (1906, p. 549) urged that it be applied to "practical statistical problems":

> The attention of mathematical statisticians has been mainly directed to theory, and to actual measurements of biological and anthropometrical correlations; it is time that it was brought to bear on the criticism and analysis of existing industrial statistics.

Edgeworth and Pearson must have been in his mind.

Regarding the method of sampling, there were no investigation to report but Bowley (1906, p. 553) had a proposal:

> It is frequently impossible to cover a whole area, as the census does, or as Mr. Rowntree here [in York] and Mr. Booth in London successfully accomplished, but it is not necessary. We can obtain as good results as we please by sampling, and very often small samples are enough; the only difficulty is to ensure that every person or thing has the same chance of inclusion in the investigation.

Life and Labour of the People in London by Charles Booth (1840-1916) and *Poverty, A Study of Town Life* by Seebohm Rowntree (1871-1954) aimed for complete enumeration. When Bowley eventually implemented his proposal he changed the way poverty was studied in Britain; see Hennock (1987) and several chapters in Bulmer et al. (1991). Bowley looked back in his history of surveys (1936) and obituary of Rowntree (1955). For other discussions of the address see Stephan (1948), Cheng (1951) and Bellhouse (1988).

Bowley (1906, pp. 550-53) illustrated the feasibility of sampling with an experiment in which a sample of 400 was taken from a list of yields for 3,878 companies; the selection was done using the final digits in a table from the Nautical Almanac. Bowley considers the 400 values both as a sample of 400 and as 40 samples of 10. In the former guise he (p. 551n) deduces that the average is "known with practical certainty to be between £4. 7s. and £5 and the chances are even that it is or is not between £4. 13s. 3d and £4. 16s. 6d. Actually when all the 3,878 were added together, the average proved to be £4. 15s. 7d." In the latter guise Bowley shows that the 40 averages "conform fairly with a normal curve of error." Looking back, it is surprising that Bowley had not made the proposal in 1901–perhaps being in York with Rowntree gave his thoughts a new focus. The section on "samples" in the *Elements* (1901, pp. 308ff.) begins by noting the ubiquity of the "method of sampling" in business and everyday life and proceeds to give examples like the one in 1906. Thus from a collection of 122 death-rates "which clearly do not conform to the normal curve" Bowley (p. 309) generates 18 random samples of size 4 the averages from which are found to "fit a curve of error closely." However, while the motions were the same as later, in 1901 the point of providing an empirical "confirmation" of the Edgeworth central limit theorem was to show that conditions were right for significance tests based on the normal distribution. In these early sampling exercises Bowley never seemed to wonder whether the conditions of the Edgeworth theorem were satisfied.

The proposal re-appeared when Bowley (1908, p. 466; also p. 494fn1) discussed ways of improving official statistics. More remarkably, his *Elementary Manual of Statistics* (1910) has a chapter on sampling based on the 1906 address. Although the method had not been used in practice, Bowley presented it with others he (1910, p. v) said were "indispensable in the handling of numbers on a large scale." Bowley (p. 63) had this comment on other ways of sampling:

> Rather than trust to the arbitrary action of chance, some investigators prefer to choose what they believe to be typical groups, and examine them in detail. Thus, investigations as to the wages, etc., of agricultural labourers have been conducted by selecting some forty districts of all kinds of agriculture, and of all economic situations. This method results in an accurate and intelligible picture, but there is no easy means of calculating any average, or of knowing the distribution by number of persons earning various rates of wages.

This appears to be his only published comment on other practices before he joined the International Statistical Institute's commission on sampling in 1924.

Other practices had been discussed at the ISI in 1895, -97, 1901 and -03; see Kruskal and Mosteller (1980). Allen & George (1957, p. 237) report that Bowley attended most of the ISI sessions after he became a member in 1903. Bowley never referred to those debates or to the practice of Anders Kiaer on which they centred. If he had studied them and Desrosières (1991, p. 236) suggests he did, he may well have discounted Kiaer's work because it made no explicit use of probability and so ignored the "test of precision" this made possible. When at last Bowley (1926a, p. 64) spoke to the ISI on sampling it was to say, "[I]f no measurement of precision can be made, the results are valueless". See Section 6 below.

In 1912 Edgeworth was president of the Statistical Society and for his address he chose the themes Bowley had used at the British Association in 1906. Edgeworth (1913, p. 175) commended sampling–"Mr. Bowley rightly anticipates much saving of trouble from the use of such methods"–but he also described Kiaer's practice and the ISI debates it generated. Edgeworth (pp. 175-6) also took up a problem that had puzzled Bowley in 1906: for one category there was a discrepancy between a population value of 7.25% and a sample value of 3.8% based on a sample of 400. Edgeworth showed that it was not so improbable when instead of looking at this most extreme discrepancy "Professor Pearson's beautiful criterion" is applied to *all* the categories. Edgeworth (p. 176) also used the Pearsonian criterion to consider the agreement between Kiaer's sample values and the corresponding census values.

Edgeworth described some of Bowley's experiments but Bowley was already doing the real thing. The first survey and the one that set the pattern was of working-class households in Reading. "The results are of much more than local interest since they prove that an inquiry adequate for many purposes can be made rapidly and inexpensively by a proper method of samples" Bowley (1913, p. 672) declared. In 1906 Bowley had used tables to generate a random sample but now he (p. 672) took one building in twenty in each street from a list of buildings. When he described his results it was with the rider, "if the conditions of random sampling are secured, as it is believed they have been in this case". Later he interpreted the practice as a form of stratified sampling.

For "theory and method" Bowley referred to the 1906 paper, where he (1913, p. 673) said,

> It is demonstrated mathematically that if in our sample 622 working-class households we find respectively 5, 10, 20, 40, 50 per cent. of cases, we may expect that the percentages in the whole are within 5 ± 1, 10 ± 1, $20 \pm 1\frac{1}{2}$, 40 ± 2, 50 ± 2, ...

The scheme "proportion $\pm$ standard deviation" was modelled on the scheme "estimate $\pm$ probable error" used in the theory of errors for the even odds interval.

Given that the probable error is .6745, roughly $\frac{2}{3}$, of the standard deviation, Bowley's "we may expect" reflects odds of about 2 to 1.

Other English towns were surveyed and the findings were collected in Bowley and Burnett-Hurst's *Livelihood and Poverty* (1915). Bowley contributed the chapter on method, "Criticism of the accuracy of the results." Bowley (1915, p. 174) considered four sources of error: incorrect information, loose definitions, bias in selection of sample and "calculable possibilities of error arising from estimating the whole by measuring a part." Allen (1968, p. 135) remarks that the account "would have been readily accepted two generations later." To show the type of conclusion that can obtained from a random sample Bowley (1915, p. 179-80) considers the case of a group of N things, pN of which have some assigned character, and n are chosen at random:

> There is, in fact, a general table of probability applicable to such a case. It is proved that it is just as likely as not–the odds are equal–that the number found in the sample will differ from pn by as much as $\frac{2}{3}$ of $\sqrt{\frac{p(1-p)}{n}}$. This quantity is called the *probable error* of the measurement. Conversely, it can be shown that (unless p is very small) if $p'n$ examples are found in n trials, it is as likely as not that the proportion in the whole group will differ from p' by as much as $\frac{2}{3}\sqrt{\frac{p'(1-p')}{n}}$. When the "probable error" is established, the tables of probability show that the

> fact will differ from the forecast only once in 25 experiments in the long run, and by four or five times this error so seldom that the chance of so great a deviation is negligible.

Bowley seems to be spinning through phases of inference: a direct probability statement based on the normal approximation to the binomial is followed by a statement of inverse probability for the proportion in the whole group and then directness is restored with "in the long run." Matters would be clearer in the 1920 revision of the *Elements*.

Bowley had been in the habit of quoting distributional results without much thought for their appropriateness but from 1920 he showed more concern with matching theory and practice. There is a sign of the new attitude in a 1918 paper by Isserlis on a finite population central limit theorem when Isserlis (1918, p. 75) reports that Bowley had drawn his attention to the importance of the problem in the theory of sampling. Leon Isserlis (1881-1966) is an interesting figure who linked several communities: he was a Pearson student and *Biometrika* author who knew Russian and championed Russian work, especially Chuprov's–see Isserlis (1926), Greenwood and Isserlis (1927) and Irwin (1966). Aleksander Aleksandrovich Chuprov (1874-1926) and others in Russia had taken up Bowley's ideas on sampling and were quicker to develop the supporting theory and arguably took it further; see Seneta (1985). After 1917 Chuprov published several articles in

English journals, in *Biometrika* (1918, -19 and -21) and in the *Statistical Journal* (1925). The writing was very technical and no articles in those journals were so un-applied yet so applicable. The last, for instance, on the law of large numbers for correlated observations covers the case of sampling from an urn without replacement (1925, p. 96). In 1923 Chuprov published a two-part paper in Gini's journal *Metron.* Chuprov (1923, p. 469) states in the introduction to the first part that the formulae are of "great practical use" for the method of sampling "which becomes in our days more and more popular" but he does not refer to anybody's work in particular. As it happens, the piece appears next to a note by Bowley–the latter is discussed in Section 5 below. The second part of Chuprov's paper contained the formula for optimal stratification that Neyman (1934) rediscovered. Seneta (1985) also describes the work of the statisticians who stayed in the Soviet Union but that work had no impact on developments in the West, certainly none on Bowley.

4 Distribution theory and inference: 1920

Bowley took his biggest single step to the memoir of 1926 when he produced a fourth edition of the *Elements* in 1920. The 200 pages of mathematical statistics in Part II constituted "practically a new work" wrote Yule (1921, p. 222). The original Part II–see Section 2 above–never reflected Bowley's immersion in the subject.

In 1902 he collaborated with Edgeworth on "Methods of representing statistics of wages and other groups not fulfilling the normal law of error" (1902). Bowley was responsible for the empirical side but he understood Edgeworth's theory and he could use Pearson's methods–the method of moments, χ^2 and the system of curves. Some material from this project was used in an appendix to the second edition of 1903 but there was no new technical material in the third edition of 1909. There was no reflection of Bowley's growing interest in sampling, though that changed too in 1920 so that Yule (1921, p. 223) could report, "Nowhere else will [the reader] find, certainly in any English work, so many illustrations of the theory of sampling as applied to social investigation."

In most respects the new Part II was the old in long trousers. In 1901 Bowley had stopped at asymptotic normality but now he went on to higher order approximations, to Edgeworth's (1906) "generalised law of error or law of great numbers." Before he had stopped at simple correlation, now he went on to multiple correlation and to association and contingency; Yule's *Introduction* was an influence here. Pearson's (1900) paper introducing χ^2 had appeared while the first edition was in press, now there was a chapter on "Tests of correspondence between data and formulae" and a discussion of χ^2 in relation to contingency. Of the new topics, the one Bowley felt strongest about was inverse probability. The original Part II had not touched inverse probability but it was always part of Edgeworth's scheme and it was in the atmosphere around the explanation Bowley gave in 1915 of the accuracy of survey results.

The standing of inverse probability had long been uncertain: there was a string of critics and those who used it did not use it exclusively; Dale (1999) gives an overview while Zabell (1989) concentrates on some of the critics and Aldrich (2007) on some of the users. Even Edgeworth presented direct as well as inverse arguments in his probable error papers (1908/9), conceding (1909, p. 82) that the direct method is "free from the speculative character which attaches to inverse probability." In the preface to the new *Elements* Bowley (1920, p. 8) declared unequivocally:

> If we are judging a universe from a sample, we have not arrived at any definite result till we can make such a statement as "the most probable average (or whatever may be the quantity in question) in the universe is A, and the chance that the average differs from A by $d_1, d_2...$ are $p_1, p_2...$"; this involves a definite use of inverse probability and of a table of probability. A mere statement of the standard deviation tells us very little.

Bowley probably had Yule and Pearson in mind for they often seemed satisfied with that "mere statement"; Aldrich (2007) describes their practice. In his review Yule (1921, p. 222) merely noted that the *Elements* has a chapter "on the very difficult subject of the measurement of precision of statistical constants from the standpoint of inverse probability."

Most of Part II is taken up with direct probability with establishing the properties of a sample drawn from a universe. For Bowley this was not a diversion from the inverse argument but a necessary preliminary, as we see below. The chapter on the "law of great numbers" goes beyond the normal approximation to the second approximation. Bowley (1920, p. 302) shows how the normal distribution function has to be modified by a skewness term based on the product of $\kappa = \mu_3/\sigma^3$ and $f(z)$ where $z = x/\sigma$ and

$$f(z) = \frac{1}{6\sqrt{2\pi}}\left(1 - (1 - z^2)e^{-\frac{1}{2}z^2}\right).$$

Bowley provides a table of values of $f(z)$ and he fits the second approximation to several sets of observations; in one case (p. 310) he fits a Pearson Type III as well. There is a brief discussion (pp. 344-6) of the Pearson curves under the heading "empirical frequency curves." Bowley does not go on "to discuss how far these equations can be used in questions of probability, nor to consider how far the fundamental formula is empirical and how far it is dependent on hypotheses of chance generation" but his opinion seems clear.

Bowley treats several aspects of sampling from a "limited universe" (finite population). He (1920, pp. 282-4) looks first at a random sample of n objects from a universe of N, pN of which have a certain quality. Having obtained the hypergeometric distribution, he then shows how it tends to normality when n is large. In the parallel investigation for measurements he (pp. 300-1) finds the standard deviation of the sample mean and quotes the Isserlis (1918) central limit theorem.

The analysis of the limited universe was not Bowley's only move towards sampling as he practised it. Stratified sampling had a role, as he (p. 332) observed:

> It may happen, however, that the universe consists of different regions or strata in which the chances are different, and the question arises whether we should proceed at random in the selection of a sample out of the universe as a whole, or whether we should partially arrange the choice so as to take the same proportion out of each region or stratum.

Bowley derives formulae for the standard deviation of the sample proportion (p. 333) and of the sample mean (p. 337). He observes that "by choosing proportionately from various strata the standard deviation of the error involved is diminished." He does not consider the possibility that an even better choice could be made by taking unequal proportions from the strata.

Bowley interpreted his own practice in Reading and elsewhere as stratified sampling: "instead of numbering all the houses and selecting 1 in 20 at random, we marked one of every 20 throughout each street." He thought the procedure had two advantages over unrestricted random sampling: “By this means we secured that no district was completely unrepresented, which may possibly happen in a random selection, and we also got the advantage given by the formula just given, since social conditions in a street have a certain similarity.” Bowley put more weight on the first consideration for his illustrative calculations showed that the standard deviation was not much diminished by his form of stratification.

The formulae came from Yule’s *Introduction*, from Part III on the “theory of sampling.” Yule (1911, pp. 281 & 345) was analysing the effects of removing the limitations of simple sampling by allowing the chance of a success to vary from trial to trial. Yule did not focus on design as Bowley did and his sampling is more often the work of nature or a gamester than the statistician. In fact, Yule appears to have been sceptical about survey sampling, at least to investigate working-class income and expenditure. He (1911, p. 276) objected that lower income families are “almost certain to be under-represented” and even if they were compelled to participate the information they would supply would be unreliable. When Bowley advocated sampling before the Society in 1908 Yule (1908, p. 487) commented that “in a practical case it was practically impossible to get a purely random sample. ” When reviewing the *Elements* he (1921, p. 223) particularly drew the reader’s attention to “the rules and cautions” pertaining to sampling given in the book (pp. 278-80).

The new *Elements* has three chapters on correlation and regression and the methods are illustrated using data from the Reading survey, thus Bowley (1920, p. 403) reports the partial correlation between rent and household size given income as $-.136$ with a standard deviation of .040. Bowley (1922) used this example when defending the procedures of statisticians against Keynes's contention in the *Treatise on Probability* (1921) that statisticians were apt to make unsupported generalisations. Bowley (1922, p. 98) insisted that "a very great part of statistical analysis is concerned not with extending a measurement made on one group to other separate groups, but in inferring from a sample selected at random or by rule from a single group what is the constitution of the group." His own inductive generalisation is to the universe of working-class households in Reading from which the sample had been taken and not to any wider group.

Bowley (1922, p. 99) also explained how the information $-.136 \pm .040$ is used

> the chance is about 2 to 1 against the observed value appearing if the value for the whole town was not between $-.10$ and $-.18$, and very great against its appearance if the value were beyond, say, $-.4$. We can make such a statement more definite if we use inverse probability, and can then estimate the probability of the value for the town.

In Bowley's discussion of correlation there seems to be an unconscious appeal to a super-population from which the Reading population is drawn for the only correlation surface Bowley knew was that associated with an infinite population; he had the concept of a proportion in an urn but not of a correlation in an urn.

After eight chapters of direct arguments Bowley reaches the inverse question– "what we can infer about a universe from a given sample?" He begins the chapter on the "Precision of measurements of averages, moments and correlations" by emphasising the importance of inverse probability and giving an example of the reasoning involved. Bowley's (1920, p. 412) paradigm involves inference to p where pN is the number of things having a certain characteristic in a universe of N things, from which n are selected at random and $p'n$ are found to have the characteristic. Bowley assumes that "all values of p from 0 to 1 are a priori equally probable." The derivation of the large sample normal approximation to the posterior follows Todhunter's (1865, p. 554) gloss on an argument from Laplace; we saw the result was hovering over Bowley (1915). The derivation is actually for the infinite universe case and Bowley (p. 414) just writes in the finite population correction at the end.

Bowley admits that the assumption of a uniform prior is a "difficulty" but there is an escape which he (p. 413) motivates by considering the case of $p' = .1$ and $n = 100$. His calculations show the likelihood–as we would call it–falling away sharply as p deviates from .1. Bowley then states:

> This example then illustrates a theorem that we may give as obvious:

> that, except in the neighbourhood of the central value, it is indifferent what distribution of *a priori* probabilities p we suppose. Over the small important central region the assumption that the *a priori* probability of p over a region is proportional to that region is likely to be a good first approximation.

Bowley refers to Edgeworth (1908, p. 387) and Edgeworth refers back to his own "À priori probabilities" where he (1884b, p. 207) had considered the case of the mean of the normal distribution. Large-sample robustness was not Edgeworth's only defence of the use of a uniform prior: in a commentary on Venn he (1884a, p. 224) argued that Venn's "logical scepticism had often carried him too far from the position held by the majority of previous writers upon Chance" and he (p. 231) went on to maintain that, "probability-constants occurring in nature present every variety of fractional value; and that natural constants in general are found to show no preference for one number rather than another." In other words, the uniform prior was in rough correspondence with experience. Pearson took this line and it long featured in his writing on Bayes–see Dale (1999, pp. 504-18) and Aldrich (2007). Bowley never did and in his two inference pieces (1923 and -26b) elaborate the large sample robustness argument. For more on Edgeworth's commitment to inverse probability see Bowley (1928, ch. II), Dale (1999, pp. 439-47) and Section 7 below.

Having presented the binomial example in some detail Bowley sketches a "general method." This large-sample distribution-free method became the template for his work on inverse probability. Writing X' for "any given function of n samples chosen at random from a population where the (unknown) corresponding function is X" and x for the difference $X - X'$, Bowley (1920, p. 415) claims:

> If we can show that the chance of obtaining the value X', when the value in the universe is X, is of the form $P_x = P_0 e^{-\frac{x^2}{\sigma^2}}$ ... then we can affirm with reasonable certainty that the sample gives evidence that the most probable value of the function in question is X' and the chance of deviations from X' is given by the normal function with standard deviation σ.

Bowley does not specify the distribution of the individual observations beyond making the implicit assumption that the large-sample distribution for X is valid. A further consideration separates Bowley from Harold Jeffreys and modern Bayesians. Bowley saw the inverse argument as completing the direct argument while Jeffreys (1939, p. 315) rejected the use of the sampling distribution in inference. Thus Bowley obtains the posterior of X, call it $p(X\,|X')$, by combining $p(X'\,|X\,)$, the sampling distribution of X', and $p(X)$, the prior for X :

$$p(X\,|X') \propto p(X'\,|X\,)p(X).$$

Bowley's strategy resembles Student's in the correlation paper of 1908: to obtain the posterior distribution of the population correlation Student first looked for the distribution of the sample correlation. The strategy presupposes a choice of statistic but neither Bowley nor Student recognised the point and just did what came naturally. In 1920 the useful idea of a sufficient statistic had not quite been born and in any case would only be applied to Bayesian inference by Jeffreys; see Aldrich (2002 and -05). The similarity between Student and Bowley does not extend very far for Student concentrated on the small-sample problem where the choice of prior would matter while Bowley was not interested in small sample problems. Bowley never referred to Student's work, not even to the direct t theory which he must have known from Fisher's writings.

Most of the results in the chapter on inverse probability were from the papers on probable errors by Pearson & Filon, Sheppard and Edgeworth; the papers are discussed in Aldrich (2007). In the main they are *not* inverse probability results but Bowley believed that out of such direct results inverse results would come.

The *Elements* left a lot to be done if all the direct probability arguments had to be 'inverted' to produce "definite" results. However before Bowley did anything there was a new element in the situation–Ronald Fisher and his rejection of the Bayesian approach. "The theory of inverse probability is founded upon an error, and must be wholly rejected" was a characteristic declaration–this from Fisher

(1925a, p. 10). Fisher had been publishing in a small way since 1912 but he only came to Bowley's notice in 1920; the course of Fisher's thinking on inverse probability is examined in Aldrich (1997).

5 Defending inverse probability: 1921-3

People were writing about inverse probability: Wrinch and Jeffreys (1919) and Pearson (1920) were writing for it, Fisher (1922) against it with Keynes (1921) somewhere in between. Bowley did not know the paper by Wrinch and Jeffreys which bore the future of inverse probability in Britain but he knew the other works. Pearson (1920) advanced the unlikely claim that the Bayes-Laplace result for the probability of future successes in Bernoulli trials did *not* depend on the assumption of a uniform prior. Edgeworth (1921, pp. 82-3fn.) jumped on this and supplied references from Cournot to the *Elements* for the valid *large-sample* argument. Bowley was only a spectator but he commented to Fisher that Pearson's error was a case of "mistaken zeal." Pearson's paper and the reactions it provoked are discussed by Aldrich (2006 and -7).

When he reviewed Keynes's book Bowley (1922, p. 99) found little that was new or telling in its reservations about inverse probability:

> The inverse step ... is admitted by all competent statisticians and mathematicians to involve assumptions of a difficult nature, and there

> is now a good deal of discussion as to the circumstances under which such assumptions are valid. ... [The uniform prior] is, if used at all, put in a modified form, viz., that nearly the same result is obtained whatever the distribution of chances of the values of P in the neighbourhood of the observed p, which is the only region concerned when the number in the sample is large.

Keynes's critique and the statisticians' response to it are discussed at length in Aldrich (2008).

In June 1920 Ronald Fisher (1890-1962) joined the Statistical Society. The biometrician had fallen out with Pearson, and was seeking other outlets for his work, see Box (1978) for biographical information. Bowley and Fisher started out by agreeing to disagree with Pearson. In the first of his papers on χ^2 Fisher (1922a, p. 91) praised the treatment of the four-fold table in the *Elements* (1920, pp. 371-2) and Bowley in turn recommended Fisher's paper to the *Journal*, "it is of considerable importance and originality"; his assessment is printed in Box (1978, p. 85). Later, however, some space between their views on χ^2 opened up–see Fisher (1923 and -4) and Bowley & Connor (1923).

A few days after commending one Fisher paper Bowley was commenting on a draft of another. "On the mathematical foundations of theoretical statistics" has been described by Stigler (2005, p. 47) as one of those rare cases when "a single work launches a new era in a science." The paper was very long, very dense and very difficult and Bowley only gave Fisher some general impressions in his letter of May 25th 1921. It begins, "The method seems powerful & in a certain region useful, It is so new (as far as I know) that I shall be very glad to see the paper published." In such an elaborate paper it is not exactly clear what Bowley was praising. However there is no mistaking his wish to defend the body of work he had presented in the *Elements*. Fisher (1922, p. 310) refers to the "prolonged neglect into which the study of statistics in its theoretical aspects, has fallen." Bowley could not accept this, "Nearly everything in Part II of my book has been developed since 1895, & I cannot admit that Part is not theoretical, even if the theory is wrong." Further on in the paper Fisher (p. 326) developed the theme of the "baseless character of the assumptions made under the titles of inverse probability and Bayes' Theorem." Bowley was surprisingly restrained, "Nor am I at all certain that a priori probabilities have not a considerable use (as on p. 414 of my book)–that the method can be abused does not differentiate it from many other methods." Bowley was setting Fisher an example in restraint for he thought Fisher should moderate his controversial tone. He finished by saying, "A difference in tone in handling other people's mistakes need not prevent sound criticism." Fisher could not have taken Bowley's advice for the comments have a prefect fit to the published article. Bowley's attitude to the paper resembled that of the referees, Yule and Eddington, as reported by Edwards (1997, p. 181). All welcomed the paper but none changed what they did, at least not in the direction of Fisher.

Bowley produced a note on "The precision of measurements based on samples" (1923) which aimed "to define certain conditions of preliminary knowledge under which inferences can be made from the known to the relatively unknown." Bowley refers to Keynes and Pearson but the note seems aimed more specifically at Fisher who is not mentioned! It is Fisher's (1922, pp. 325-6) challenge to the uniform prior that Bowley (p. 494) paraphrases:

> the hypothesis that every value from 0 to 1 is equally probable is not only baseless but is also inconsistent with an equally plausible hypothesis that all values of *arcsin* r from 0 to 1 are equally probable.

To save inverse probability Bowley elaborated the large-sample robustness argument sketched in the *Elements*.

Bowley presents three cases, inference to the probability of a success in Bernoulli trials, to the mean of a general distribution and to an arbitrary parameter. The first was treated in the *Elements* and also in Fisher (1922, pp. 324-6) where the Bayes' theorem and maximum likelihood treatments are compared. Bowley (1923, p. 495) sets it up, "From an infinite universe n things are taken independently and at random, and of them pn are found to possess a certain attribute." He assumes that $F(p')$, the prior for p' the unknown proportion in the universe, to be "expansible by Taylor's Theorem" and to have finite derivatives and uses an Edgeworth approximation for the density of pn. He shows that if terms of order $1/n$ are neglected that the posterior takes the classical Laplace-Todhunter normal form–"whatever the form of F." One of the curious features of all Bowley's 'Edgeworth meets Bayes' examples–both here and subsequently in the memorandum–is that the limiting posterior is normal and so his derivations are sophisticated rationalisations of doing what comes naturally.

The treatment of the second case is more like a special case of the "general method" from the *Elements* but with a higher order approximation replacing the normal approximation. The inference is from $\overline{x}$ the mean of a random sample to $\overline{x}'$ the mean of the population from which the sample is drawn. Bowley (1923, p. 497) does not specify a form for the population but assumes that there is an Edgeworth expansion for the distribution of $\overline{x}$. Having made a similar assumption about the prior $F(\overline{x}')$ as in the binomial case, he reaches a similar conclusion, viz. that "the chance that the average in the universe was within the limits $\overline{x} \pm x$" can be calculated from the normal distribution. Bowley recognises that σ, the "standard deviation of the magnitudes in the universe," is "not generally known" and makes an allowance for this in the asymptotic analysis though without achieving the desired approximation. In the third analysis (p. 498) the parameter is the population counterpart of a linear function of the observations. Bowley thought that this cleared up the remaining important cases, higher moments and correlation, but his scheme does not actually cover the correlation case.

By supplying these rather weak “conditions of preliminary knowledge under which inferences can be made from the known to the relatively unknown” Bowley responded to Fisher’s nihilism. This veiled response was the only response to Fisher (1922) from the senior statisticians in the 1920s. There is no indication that Fisher even noticed the note. He and Bowley went on playing bridge together–they were neighbours–and in 1930 Bowley suggested that Fisher join the ISI but they stopped discussing each other’s work. The work went in different directions. Fisher (1925a, -b) elaborated his ideas on the theory of estimation, worked on Student’s distribution and developed his fundamental ideas on the design and analysis of experiments while Bowley’s activities in statistical theory were confined to sampling and to putting Edgeworth contributions into order. The latter project–Bowley (1928a)–was his last big piece on mathematical statistics and it was more of a historical afterword to the *Elements* than a platform for future work. Later Bowley (1935, p. 56) registered a “similarity of part of Fisher’s work [on the efficiency of maximum likelihood estimation] to that of Edgeworth” but Bowley had never wanted to develop that common part himself; the similarity claim is discussed by Pratt (1976). The paths of Bowley and Fisher were not so entirely distinct for Fisher, in effect, created a sampling school at Rothamsted; see Section 8 below.

After the *Elements* and the note came Bowley's most ambitious work of statistical theory, the "Measurement of the precision attained in sampling" written for the International Statistical Institute. It contained two theories and, while the theory of random selection was true to type, the theory of purposive selection was a sport.

6 Sampling theory 1924/6

At the 1923 congress of the ISI Adolph Jensen (1866-1948), director of the Danish Central Statistical Bureau, presented a paper and after a 20 year hiatus sampling was back on the agenda. Jensen (1923) reviewed the earlier discussions and reported some new studies; in his (1926b) survey of surveys these appear under the headings "random selection of groups" (pp. 394-5) and "purposive selection of groups" (pp. 408-411). In May 1924 a commission was appointed to investigate the "application of the representative method." Its most active members would be Jensen and Bowley; Jensen was responsible for the report to which they both contributed long annexes.

In June 1924 Bowley commented on the task facing the commission. The occasion was noteworthy for marking a small step towards official adoption of

Bowley's method: John Hilton, Director of Statistics at the Ministry of Labour, was describing a successful experiment in sampling; see Stephan (1948, p. 22) and Moser (1949, p. 232). In the discussion Bowley (1924, p. 564) referred to the purposive method and the challenge of clarifying its mathematics:

> I have no doubt that method may yield very good results, and Dr. Jensen was able to show the same kind of agreement as Mr. Hilton has shown today, but the mathematics of the method is obscure. We must not neglect that method; it must be studied and some measure of precision found for it.

Yule and Edgeworth were also discussants: as usual Yule (1924) complained about the difficulties of getting a random sample while Edgeworth (1924) welcomed the advance of sampling though he had some reservations about Hilton's treatment.

The commission presented its findings at the Rome congress in September 1925; Kruskal and Mosteller (1980, pp. 185-7) and Cheng (1951, pp. 222-3) describe the meeting and the resolutions. The report–Jensen (1926a)–describes two forms of the representative method, "random selection" and "purposive selection." The first was Bowley's method but it was recognisable–at least as a theoretical possibility–to anyone with a knowledge of probability. Jensen (1926a, p. 361) characterised the second as

> the method of selecting a number of groups of units in such a way

> that the selected groups together yield as nearly as possible the same averages or proportions as the totality in respect of those characteristics which are already a matter of statistical knowledge.

The groups of units are chosen because they are known to be representative of the population in some respect; each group is an "approximate miniature of the population" as Kruskal and Mosteller (1980, p. 175) put it. From Jensen's (1926b) Annexe B on the representative method in practice it appears that the purposive method was much more widely used than the method of random selection. As to the merits of the methods, Jensen (1926a, p. 376) stressed the practicality of the purposive method, while conceding that "the precision can generally not be so easily measured by mathematical means." Bowley's surveys were all of a pattern, with a list of units from which the sample was chosen and he never faced the situations that led others to purposive sampling.

There may have been something of Bowley in the report but his main effort was Annexe A, "The precision attained in sampling"–60 pages of formulae, derivations and numerical examples. At the meeting Bowley spoke in more general terms, as when he (1926a, p. 63) explained why the analysis was so important:

> It is completely unscientific to say that a percentage is "about 30"; we must be able to assign reasons for believing that it is "between 28 and 32". The grounds for such a belief are based on the mathematical

> theory of probability, and it is to application of this theory to the problem that my analysis is devoted.

He (1926a, p. 64) ended by re-iterating how "the problem of representative methods is primarily one of mathematics."

While random selection was already mathematics, purposive selection had yet to be treated mathematically. Bowley (1926a, p. 63) recalled how he had viewed the prospect:

> for a long time I thought that no mathematical measure [of precision] could be given, and that we must depend either on good judgement or on repetition of observation. But I presently discovered that the mathematics of correlation gave powerful assistance.

We glimpsed something of that old attitude in Section 3 above.

Bowley's annexe treats two forms of the representative method, random selection and purposive selection, with the aim in each case of measuring the "precision attained in sampling." While Jensen's annexe on practice was a guide to the literature, Bowley's was a guide to his own mind, to the theory as it should be. The annexe has two parts with the longer first part on random selection. Here Bowley (1926b, p. 22) acknowledged the existence of results from "the time of Laplace, Gauss, Bernoulli and Poisson" and others from "long ago" but gave no references. His interest was in extending those results.

The extensions were in predictable directions. On the side of direct analysis that Bowley (1926b, p. 22) saw the need as follows:

> so far as I can ascertain, no one has brought together these formulae so as to give the frequency correct to the second (or $1 \div \sqrt{n}$) term, when the universe is restricted, or when the sample is stratified, or when both these conditions apply, either for variables or for attributes. To obtain these frequency curves, it is necessary to go back to first principles, and in the following pages the lines of well-known analysis are developed.

Those "first principles" and the "lines of well-known analysis" were well-known from the *Elements*. Bowley provides no references for orientation, only references for technique–to Edgeworth (1906) for second-order approximations, Pearson (1900) for χ^2 and Bowley (1923) for the Bayesian argument–and for the occasional algebraic detail–to Pearson (1903) and Chuprov (1921).

Bowley (1926b, p. 23) also identified a need on the side of inverse probability:

> Most of the work will be recognized as a simple extension of generally accepted principles; the problem before us is definitely to make inferences from a given sample to an unknown universe, whereas the great bulk of recent work has proceeded from an unknown universe to

a sample and we are therefore obliged to go on to the doubtful ground of inverse probability.

Doubt and difficulty were no strangers for Bowley (1926b, p. 5) had a way:

> Emphasis is laid on the logical difficulty of making a quantitative inference from the sample to the universe ... and a method is suggested by which this difficulty can be overcome.

The method came from the 1923 note.

Part I of the annexe (pp. 6-45) is a comprehensive treatment of random selection presenting direct and inverse results for sampling with and without stratification for a single attribute, alternative attributes and a single variable. Bowley poured content into all the boxes indicated in the *Elements* plus the extra box of alternative attributes to which Edgeworth had drawn attention in 1912. Or from another point of view, he analysed the restricted universe analogues of the cases he had treated in the Bayesian note of 1923.

Much of Part I is taken up describing and extending the results of Edgeworth (1906) to the case of sampling from a restricted universe. The least involved of Bowley's cases is the single attribute for which he (p. 40) develops an approximate expression for the density of the sample proportion

$$E_x = \frac{1}{\sigma'\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma'^2}} \left\{ 1 - (q-p)\left(1 - \frac{1}{2}k\right)\frac{x^2}{2\sigma'^4} \right\}$$

where $\sigma'^2 = pqn(1-k)$ and $n = kN$ with n the sample size and N the universe size. He then (p. 43) inserts this into Bayes' formula and proceeds as in the 1923 note. Hald (1998, pp. 292-4) provides a commentary on the derivation.

Part I brought together statistical theory as Bowley understood it with sampling as he practised it and completed the development begun in 1906. Part II was a new project, applying the general statistical theory of the *Elements* to a different form of sampling, purposive sampling. Here there were no results or even standard practice to go on and the only publication Bowley mentions is Jensen (1923) and that only incidentally. In purposive sampling information on other characteristics is used and Bowley thought the mechanism had to be understood in terms of correlation and so correlation became part of his (1926b, p. 5) specification of purposive selection:

> Here the unit of selection is a district or group, every member of which is included in the sample. The selection is made so that the aggregate of the districts gives the same results as the universe in respect of certain quantities (called "controls") which are known in the districts and universe and which are correlated with the unknown proportions or quantities which are the subject of investigation.

The representativeness that Jensen emphasised had only a supporting role in Bowley's theory where the main work was done by the correlation between the known and the unknown.

Bowley (1926b, p. 46) starts by declaring that the method of purposive selection presents problems which "differ in emphasis, rather than in kind" from those presented by random selection and that the "essential difference" lies in what is being selected, a district versus an individual. This is an arresting observation for in the theory of random selection the measure of precision rests on knowing the chance of selection while in purposive selection there is no calculable chance of selection: apart from representativeness on the controls, "the selection followed no rule" he (p. 50) says of one of his examples. For the modern reader Parts I and II are chalk and cheese–in modern terms, I a form of design-based inference using Bayesian principles and II a form of model-based inference using direct probability–but Bowley moves between them without remark. Given Bowley's history and his answer to Keynes–see above Section 4–it might have been expected that he would insist on design-based inference as a matter of principle. However, measurable precision was the goal and both the old random selection theory and the new purposive selection correlation theory used probability to reach that goal.

In the simplest set-up (pp. 47-52) there is a variable and a single control and no stratification The population of A units is distributed into N districts with the

s-th district containing a_s units so that

$$A = \sum_1^N a_s.$$

The object is to find X the average of the variable associated with every unit. If the average in a district is x_s then

$$X = \frac{\sum_1^N a_s x_s}{\sum_1^N a_s}.$$

For a sample of n districts the average value X_n is given by

$$X_n = \frac{\sum_1^n a_s x_s}{\sum_1^n a_s}.$$

Purposive sampling exploits the existence of an allied measurement whose magnitude is already known for each district. A district s is included in the sample only if its average u_s, say, matches the population average U. Given such a sample of districts it is deemed reasonable to estimate X by X_n.

While Bowley's avowed aim was to produce a measure of precision for X_n, that is not quite what he did. He gives the measure of precision for a corrected version of X_n where the correction is based on the values of the controls, the $u's$.

Bowley's (1926b, p. 47) use of correlation is based on regarding the N values of the x's and the u's as "frequency groups" with means $\overline{x}$ and $\overline{u}$ respectively and standard deviations σ_x and σ_u with a linear regression relationship between them. The formulae for the regression estimate and its standard error are based on the standard formulae given in Yule's *Introduction.* While the general idea of obtaining a more precise estimate of X using information on the allied measurements is clear, the details are obscure and, like Neyman (1934, p. 573), "I could not exactly follow the method proposed by Professor Bowley"; in the discussion Bowley (1934, p.) admitted that "the original passage is obscure."

Bowley's set-up for purposive sampling is more complicated than for random selection and the theory does not go beyond direct arguments and first approximations. Bowley does not compare the two forms of selection or even discuss the circumstances in which the purposive method should be used; perhaps that was left for Jensen's report. The point Bowley (1926a, p. 63; cf. 1926b, p. 5) most wanted to make about the purposive method was not about the method *per se* but about the value of additional controls: "the importance of these controls has been greatly exaggerated, that it is rarely that they can increase the precision by as much as 30%, and that it is useless to increase their number. This is contrary to the opinions that have been formed on non-mathematical grounds." In the report Jensen (1926a, p. 272) quoted Bowley's finding though he (1928) subsequently disputed it arguing that Bowley had drawn the wrong conclusion from his own formulae.

Bowley's work on purposive selection is puzzling for it was so out of character, one might say, for his breakthrough here was such a departure from everything he had done before. Bowley seemed proud of turning the method into mathematics but it never became *his* method–he never used it in his own surveys and he was not eager to defend it; see Section 8 below the account of the 1934 meeting. His method was the method of random selection with stratification. For other accounts of the memorandum see Chang (1976, pp. 304-6), Kruskal and Mosteller (1980, pp. 185-7) and Hald (1998, pp. 291-2).

Away from his theorising Bowley had been involved in surveys. In 1925 Bowley and Hogg published a sequel to *Livelihood and Poverty* in which the same towns were resurveyed. In the note on the method of sampling Bowley (1925) updated the old treatment by including the finite population correction in the formula for the standard deviation. In 1927 he began work on his biggest sampling project, *The New Survey of London Life and Labour*–Booth for a new age and with the new technology of sampling. The results appeared in nine volumes edited by H. Llewellyn Smith (1930-5). The scale was new but the method was not. Bowley (1932a, p. 31) describes two methods for selecting units, "both scientifically valid." The first is to number all the units listed and select for investigation those whose numbers are found on an independent list of random numbers. The second is to select every fiftieth unit in the list: "If care is taken that the initial number (1 or 2) is not such as to give bias (or if it is chosen at random), either method gives each unit an equal chance *a priori* of being included." The second method was used. The advantages were essentially those Bowley had given in the *Elements*; see Section 4 above. An appendix–Bowley (1932b, pp. 439ff)–explains how to calculate standard errors for groups of boroughs as well as individual boroughs; the treatment is more elaborate than before but did not involve any new principles. Bowely also used different sampling proportions in different regions.

7 Digression: Pearson's "fundamental problem"

In 1928 Karl Pearson published a Bayesian treatment of the problem of inferring the composition of a finite population from a sample. Pearson's treatment of this "very fundamental" problem in statistical practice is worth a glance because it reflects such different concerns from Bowley's. Pearson had quite an intricate relationship with inverse probability but we will not need to go very far into it; for details see Dale (1999, pp. 504-518) and Aldrich (2007).

Pearson (1928, p. 163) did not know of "anyone but Laplace who has attempted the present problem on the basis of inverse probabilities." Pearson was researching Laplace–cf. Pearson (1929)–but his paper picks up themes from his own earlier work. Hald's (1998, pp. 289-302) great commentary on Laplace includes a section, "From Laplace to Bowley (1926), Pearson (1928), and Neyman (1934)." The four divide into two apparently independent pairs, Pearson writing to improve upon Laplace and Neyman to improve upon Bowley. Pearson knew something of the *Elements* but the *Biometrika* universe did not extend to social surveys or the ISI. Nor, it seems did Bowley ever tune into Pearson's Bayesian frequency.

Pearson and Bowley started from the same place, a uniform prior and a large-sample normal approximation to the posterior; see Section 4 above. Pearson's theme was that better approximations have to be found to the exact small sample posterior, Bowley's theme was that the restriction to the uniform prior can be relaxed in large sample analysis and that better large sample approximations can be used. While their priorities conflicted, their results may be thought complementary but nobody before Hald seems to have to have considered the two together.

Pearson (1907 & -20) had previously considered the infinite population problem: an event has occurred p times on $p + q = n$ trials, what is the probability of its occurring r times in a further $r + s = m$ trials? Pearson called this "the fundamental problem of practical statistics" and it had been treated by Laplace who gave an exact formula based on the assumption of the "equal distribution of ignorance." The formula involving beta functions was not immediately useful but

Laplace also produced a normal approximation. Pearson (1907) showed that the approximation was poor except when past experience is very extensive compared with the second sample and the proportion in the population neither very small nor very large. In the *Tables for Statisticians* (1914) Pearson gave exact values for the probability for the kind of values found "in laboratory work or in the treatment of rare diseases." The 1920 paper had a double objective, to demonstrate the wider applicability of the uniform prior–see Section 5 above–and to provide more extensive tables based on better approximations to the beta integrals.

In 1928 Pearson investigated a related finite population problem: what is the distribution of p the number of individuals "marked by a special characteristic" in a population of N individuals given that r are found marked in a random sample of n? Pearson (1928) combined the hypergeometric distribution for r and a uniform prior for p to obtain $C_{p,r}$, the probability of p in the population given r in the sample. Pearson calculates the mean, mode and standard deviation and compares them with those given by the normal approximation. He then considers the use of a Pearson Type I distribution as an approximation, taking three cases: $N = 1000$, $n = 100$, $r = 20$; $N = 100$, $n = 40$, $r = 10$; $N = 30$, $n = 10$, $r = 2$. Bowley was used to working with figures like $N = 16,000$ and $n = 600$.

Pearson's main objective in the fundamental problems papers was to find an accurate way of calculating the posterior distribution given a uniform prior. Over the years Pearson changed his mind about the basis for such a prior. He began from the clear position that the uniform prior reflected general experience–see Section 4 above–but then in 1920 he convinced himself that *whatever* the nature of prior experience a uniform prior could represent it–Section 5 above–and finally in 1928 he went into total obscurity with the paper's "note on the theory of inverse probabilities" (pp. 163-6).

Pearson's paper seems to have gone unnoticed in the sampling literature–or any other–until the modern revival of Bayesian sampling theory. In his "Subjective Bayesian models in sampling finite populations" Ericson (1969) noted Pearson's paper among the earlier contributions and describes its results as a special case of his own. Those results are the expression for $C_{p,r}$ as Ericson had no interest in Pearson's further analysis. Ericson does not refer to Bowley's memorandum but then Bowley's analysis was not a special case of Ericson's.

8 Bowley, Neyman and Fisher: 1934

I began with quotations from Neyman's "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," a contribution now seen as "fundamental," "classic," "revolutionary;" "a watershed," etc.; for accounts see Smith (1976, pp. 184-5), Bellhouse (1988, pp. 7-9) and Fienberg and Tanur (1996, pp. 244-7). However, when Neyman spoke to the Society in June 1934 on the mathematical statistics of the social survey this could not have seemed a promising topic, Bowley had never raised any more than polite interest and Neyman seemed set to have an audience of one and he none too sympathetic.

Jerzy Neyman (1894-1981) had been working in statistics for a decade or so and in that time his ideas had changed considerably; see Reid (1982) for biographical information and Fienberg and Tanur (1996) for Neyman's early work on sampling. Neyman had made his English debut with a paper in *Biometrika* on sampling from a finite population, a translation of a Polish publication from 1923. The paper, Neyman (1925), on sampling from a finite population resembled work by Chuprov and Isserlis and Neyman (and Karl Pearson his editor) were rebuked by Greenwood and Isserlis (1927) for not acknowledging this. In 1934 Neyman did not dwell on this history beyond remarking (p. 561) that "random sampling has been discussed probably by more than a hundred authors." Since 1927, when he started to work with Egon Pearson, he had been breathing new air–Fisher and Student on testing and Fisher on likelihood. Neyman also played with the large sample Bayesian approach although his inspiration came from von Mises (1919) rather than from Edgeworth. The Neyman-Pearson testing project first proceeded along a double front with Neyman and Pearson (1928) taking a direct approach and Neyman (1929) an inverse approach. However with their 1933 paper on "probabilities a priori" the Bayesian development came to an end and the Neyman of 1934 was a reformed man.

Neyman's 1934 paper was "chockablock full of important ideas and insights"–cf. Kruskal and Mosteller's (1980, p. 187)–but the headline (1934, p. 588) was that the method of purposive selection was unsatisfactory and that "the only method which can be advised for general use is the method of stratified random sampling." Bowley had an interest in this conclusion and in Neyman's alternative to the Bayesian argument. Fisher, also present, had an interest for underlying Neyman's conception of what makes a sampling method satisfactory was the notion of a "confidence interval" which Neyman presented as a variant of Fisher's (1930) fiducial limits. There was a third interested party Chuprov's student, Oskar Anderson, who sent comments from Bulgaria.

The theory of purposive sampling that Neyman criticised was not Bowley's theory but Gini and Galvani's (1929) version of it. Neyman further reconstructed the theory inserting the notion of random selection: "there is no room for probabilities, standard errors, etc., where there is no random variation or random sampling" he (1934, p. 572n) wrote. Even in Neyman's account this fundamental point of principle is not emphasised but appears in a footnote. Having produced a workable version of purposive sampling Neyman (p. 586) criticised the method for resting on hypotheses which it is impossible to test except by an extensive enquiry. He argued further:

> If these hypotheses are not satisfied, which I think is rather a general case, we are not able to appreciate the accuracy of the results obtained. Thus it is not what I should call a representative method. Of course it may sometimes give perfect results, but these will be due rather to the uncontrollable intuition of the investigator and good luck than to the method itself.

As well as extensive theoretical analysis and computations Neyman's critique rested on the concepts of "a representative method of sampling" and "a consistent method of estimation" which he (1934, p. 585) explained as follows:

> I should use these words with regard to the method of sampling and to the method of estimation, if they make possible an estimate of the accuracy of the results obtained in the sense of the new form of the problem of estimation, *irrespectively of the unknown properties of the population studied.* Thus, if we are interested in a collective character X of the population π and use methods of sampling and estimation, allowing us to ascribe to every possible sample, Σ, a confidence interval $X_1(\Sigma), X_2(\Sigma)$ such that the frequency of errors in the statements
>
> $$X_1(\Sigma) \leq X \leq X_2(\Sigma)$$

> does not exceed the limit $1 - \varepsilon$ prescribed in advance, *whatever the unknown properties of the population*, I should call the method of sampling representative and the method of estimation consistent.

The procedure of Gini and Galvani did not satisfy the italicised condition for it was "consistent" only in special circumstances.

Bowley (1934b, p. 607) opened the discussion by welcoming a contribution to the great cause, "This paper will, when it is thoroughly studied, do very much to remove any remaining doubt that the mathematical approach is of fundamental importance." In emphasising methods that "make possible an estimate of the accuracy of the results" Neyman was reinforcing Bowley's oldest message. However, having related some of his own sampling experience, Bowley (p. 608) changed tone: "I feel it my duty to criticize the theory of probabilities in Section II, part 1." "The theory of probabilities a posteriori and the work of R. A. Fisher" (pp. 561-3) expounded confidence intervals and presented them as a way round what Bowley always admitted was the difficulty of the inverse method. As Neyman (p. 562) pointed out, "The whole procedure consists really in solving the problems which Professor Bowley termed direct problems." To tackle these direct problems Neyman (pp. 563-7) introduced new tools and he had a new set of reference points–Markov's theorem on least squares, Student's t distribution and E. S. Pearson's study of the robustness of inferences based on t. These were not Bowley's reference points and his reference points, Edgeworth and the law of great numbers, had no place in the new scheme. The new theory of random selection had variable sampling proportions across strata and a theory of optimal allocation but otherwise it was like a return to the status quo ex 1920. Edgeworth expansions only returned to sampling from a finite population much later; see the references in Sugden et al. (2000).

Bowley (1934b, p. 609) said that he had read Neyman's elucidation of confidence limits with great care but admitted, "I am not at sure that the 'confidence' is not a 'confidence trick.'" Neyman may have thought he was helping Bowley out of a difficulty but, having paraphrased Neyman's argument, Bowley asked:

> Do we know more than was known to Todhunter? Does it take us beyond Karl Pearson and Edgeworth? Does it really lead us towards what we need–the chance that in the universe which we are sampling the proportion is within these certain limits? I think it does not. I think we are in the position of knowing that either an improbable event has occurred or the proportion is within the limits. To balance these things we must make an estimate and form a judgement as to the likelihood of the proportion in the universe–the very thing that is supposed to be eliminated.

In the first *Elements*–Section 2 above–Bowley had not been concerned with balancing "these things" but from 1920 he had been emphasising the importance of balancing and how this could only be done using inverse probability. Neyman (1934, pp. 623-5) replied to Bowley at length emphasising how the problem had been changed and how the focus was on a "rule of behaviour" and how the probability statements "concern the results of our behaviour."

Neyman's headline, the defectiveness of purposive sampling, did not much interest Bowley and when Neyman came to reply he (1934, p. 621) expressed surprise that only Anderson wanted to defend the method. Neyman (p. 559) had thought that Bowley had treated the two methods "as it were on equal terms, as being equally to be recommended" but now he (p. 621) learnt that Bowley had mistrusted the method even when he prepared his report! In the discussion Bowley (pp. 609-10) said little about the purposive method or about his theory: "it was in some way pioneer work and I should have been astonished if no improvement should have been made in the course of time." Neyman's surprise was perfectly reasonable for in the Rome documents the doubts about the purposive method were in Jensen's (1926a) report, not in Bowley's annexe.

Fisher made an interesting contribution to the meeting. He commented on both aspects of Neyman's paper, inference and sampling. Fisher's comments on inference, and especially on the relationship between fiducial inference and confidence intervals, have attracted plenty of attention; see Zabell (1989) and Aldrich

(2000) for discussion and references. At the time Neyman (1934, p. 593fn.) saw his and Fisher's theories running "on somewhat different lines" but to the same destination. The differences between Fisher and Neyman and whether they were small or large did not matter much to Bowley as he objected to what they had in common. In the present context Fisher's remarks about sampling are more interesting and these have not attracted the same attention.

Neyman had praised Bowley's contribution but Fisher (1934, p. 614) ignored it and spoke about sampling in agricultural research:

> It was, indeed in Experimental Agriculture that an adequate technique, bringing out the different aspects of the sampling problem, and displaying comprehensively exactly how these different aspects were interrelated, was first developed.

Fisher went on to wonder whether Neyman had been influenced by his "personal experience" in Agricultural Science–high praise! Fisher then recounted "having the pleasure of collaborating with a succession of brilliant plant biologists, under whom, and especially under Drs. Maskell and Clapham, the technique was gradually perfected."

In 1928 A. R. Clapham (1904-1990) and newly appointed to Rothamsted started work on sampling crop yields. His papers, beginning with Clapham (1929), are so saturated with Fisher's ideas and methods that it is hard not to see him as the original force behind what became the Rothamsted survey school. One intriguing aspect of the Rothamsted story is the part played by J. A. Hubback (1878-1968) of the Indian Civil Service. Hubback sent his “Sampling for rice yield in Bihar and Orissa” (1927) to Fisher in January 1928 explaining how he had picked up his knowledge from textbooks and *Biometrika* publications: the textbook he quoted in his paper was Bowley’s *Elements*. In 1945 Fisher wrote that Hubback’s work had “influenced greatly the development of my methods at Rothamsted” (quoted by Mahalonobis (1946)); see Rao (2005). Of course Fisher did not need to have the possibility of randomisation pointed out to him for he (1925a) had already advocated it in agricultural experiments but it would be nice to think that experimental agriculture owed something to social surveys.

Bowley, Neyman and Fisher were re-united later in 1934. In December Fisher spoke to the Society on “The logic of inductive inference” and again Bowley proposed the vote of thanks. He (1935, p. 55) began by thanking Fisher for “his contributions to statistics in general” but then moved to criticism. Inevitably the question of direct versus inverse inference came up–“I must confess that the new [direct] method appears to me to tell us only one-half of what we really need ...”– but Bowley passed to other matters including that of the “similarity” between Fisher and Edgeworth mentioned in Section 5 above. The discussion moved Fisher to fury and his (1935, p. 76) reply began, “The acerbity, to use no stronger term, with which the customary vote of thanks has been moved and seconded [by Isserlis] ... does not I confess surprise me." To Fisher it seemed that the discussants appeared in order of "diminishing animosity." Neyman spoke last and Fisher (p. 82) was really quite friendly towards him:

> It has been, naturally, of great interest to me to follow the attempts which Drs. Neyman and Pearson have made to develop a theory of estimation independently of some of the concepts I have used. That, whenever unequivocal results have been obtained by both methods they have been identical is, of course, a gratifying of the hope that we are working along sound lines.

This may have been the last good word Fisher had for Neyman for shortly afterwards relations between them broke down.

Bowley retired from the School in 1936 though he went on writing until his death in 1957. He last wrote about statistical inference in the sixth edition of the *Elements* of 1937. For this final revision Bowley added thirteen short supplements to an unchanged text. None reflected his sampling work and only one on the goodness of fit test–see Section 5 above–recorded his own research. His final thoughts on inference appear in Supplement XII on the "method of confidence belts." Bowley (1937, p. 490) begins by observing:

> Many statisticians wish to be independent of any hypothesis about $\acute{a}$

> *priori* probability, when they draw inferences from the results of sampling. The method of confidence belts or fiduciary limits, introduced by Prof. R. A. Fisher, has this purpose and we proceed to discuss one aspect of it in the simplest case.

That simplest case was the binomial distribution and Bowley followed the treatment of Clopper and Pearson (1934). Bowley (p. 492) insists that "if we have only one sample, such as was obtained in the *New Survey of London Life and Labour*, any action (such as the supply of milk to children) must be based on that one." In such a case "any confidence we may have in our results" rests on the familiar argument that the sample is large and that the appropriate prior is available. On the latter point

> The main assumption ... is that the frequency of the occurrence of "urns" with proportions differing significantly from the central region, is not overwhelmingly great,and this may in many cases be known from general experience of the "populations" with which we are concerned.

9 Sequels

The method of random selection Bowley pioneered came eventually to be widely used. While Neyman (1934) endorsed the method and declared Bowley's memoir an achievement, he did not give his audience much reason to look at it. The Bowley approach to inference was rejected by all those making the running in British statistics–Fisher and Neyman and their followers–and the theory of the memoir died. What did pass from Bowley to Neyman and the sampling literature was the idea that there should be a theory of the method and the notion that probability sampling made a "test of precision" possible. Turning to inverse theory, this was carried forward by Harold Jeffreys but he does not mention Bowley in his compendium of Bayesian theory, the *Theory of Probability* (1939), and mentions Edgeworth only because Pearson adopted his view that the uniform prior is based on experience; see Aldrich (2002 and -5) for what Jeffreys read of the statistical literature.

Turning from ideas to people, Bowley trained no inference specialists or sampling theorists–his collaborators were social investigators, like Burnett-Hurst and Hogg. In the 1920s two more statisticians joined the School, E. C. Rhodes and R. G. D. Allen. Rhodes had worked for Pearson and produced some theory of an Edgeworthian cast but his involvement in sampling went no further than commenting on Bowley's memorandum. Allen reflected Bowley's economic side and applied mathematics to economic theory and then worked in economic statistics. When interest in surveys revived at the School–after the appointment of M. G. Kendall to a second chair in 1949–it was a fresh start. At University College Neyman did not create a sampling school that survived his departure to the United States in 1938. The only Britain centre treating the method of sampling as primarily one of mathematics was Rothamsted. The school there survived Fisher's departure and the statisticians, including Wishart (from 1929), Yates (from 1935) and Cochran (from 1936) followed the biologists.

Turning from Bowley as a contemporary to Bowley in history, how has he been remembered? The first histories appeared in the 1940s when there was enough sense of achievement to believe that there was a field worth recording. In Stephan's (1948) encyclopaedic history Bowley had his place(s)–for his (1906), (1913) and (1926b). Two reviews with a strong historical element offer contrasting perspectives. In Yates's (1946) "Review of recent statistical developments," written from Rothamsted, the modern era begins with Clapham (1929) with earlier work belonging to an undifferentiated pre-history. Mahalanobis was also an admirer of Fisher but in his history (1944, pp. 443ff) the modern era begins with the ISI discussion of 1925–with Bowley.

Bowley died in 1957 twenty years after he last wrote on statistical theory. By 1957 the modern debate on the foundations of survey sampling was under way and the Bayesian revival beginning; see Smith (1976) and Bellhouse (1988) for the debate and Fienberg (2005) for the revival. Out of those debates came some useful distinctions, design-based versus model-based inference and Bayesian versus classical inference. In terms of these categories Bowley was everywhere or nowhere. Most of the time he was a design-based Bayesian–probability entered through randomisation and statistical inference rested on inverse probability–but his version of inverse theory involved a lot of preliminary direct theory and his analysis of purposive sampling involved model-based inference using direct theory. Bowley did not have much to say about probability but his Bayesianism seems a world away from the subjectivism of Savage or Ericson.

10 References

Aldrich, J. (1992) Probability and Depreciation: A History of the Stochastic Approach to Index Numbers, *History of Political Economy*, **24**, 657-686.

_________ (1997) R. A. Fisher and the Making of Maximum Likelihood 1912-22, *Statistical Science*, **12** (3), 162-176.

_________ (2000) Fisher's "Inverse Probability" of 1930, *International Statistical Review*, **68** (2), 155-172.

_________ (2002). How Likelihood and Identification went Bayesian, *International Statistical Review*, **70** (1), 79-98.

_________ (2005) The Statistical Education of Harold Jeffreys, *International Statistical Review*, **73** (3), 289-308.

_________ (2006) Burnside's Encounters with the "Modern Theory of Statistics", Forthcoming in *Archive for Exact Sciences*.

_________ (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.

_________ (2008) Keynes among the Statisticians, *History of Political Economy*,**40** (2), 265-316.

Allen, R. G. D. (1968) Arthur Lyon Bowley. In D. L. Shils (ed.) *International Encyclopedia of the Social Sciences*, vol. 2, pp. 134-137. New York: Macmillan & Free Press.

Allen, R. D. G. and R. F. George (1957) Professor Sir Arthur Lyon Bowley: November 6th, 1869-January 21st, 1957, *Journal of the Royal Statistical Society A*, **102 (**2), 236-241.

Bellhouse, D. R. (1988) A Brief History of Random Sampling Methods. In P.R. Krishnaiah and C.R. Rao, editors, *Handbook of Statistics*, volume 6, pages 1–14. New York: Elsevier.

Booth, C. (1889-97) *Life and Labour of the People in London*, 9 volumes. London: Macmillan.

Bowley, A. H. (1972) *A Memoir of Professor Sir Arthur Bowley 1869-1957 and his Family*, Petworth Sussex: Pennygate.

Bowley, A. L. (1895) Changes in Average Wages (Nominal and Real) in the United

Kingdom between 1860 and 1891, *Journal of the Royal Statistical Society*, **58**, 223-278.

_________ (1897) Relations between the Accuracy of an Average and that of its Constituent Parts, *Journal of the Royal Statistical Society*, **60**, 855-866.

_________ (1901) *Elements of Statistics* (with substantially unaltered second and third editions in 1903 and 1909), London: King.

_________ (1906) Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science, York, 1906, *Journal of the Royal Statistical Society*, **69** (3), 540-558.

_________ (1908) The Improvement of Official Statistics, *Journal of the Royal Statistical Society*, **71** (3), 459-495.

_________ (1910) *An Elementary Manual of Statistics*, London: King and MacDonald & Evans.

_________ (1911a) The Measurement of the Accuracy of an Average, *Journal of the Royal Statistical Society*, **75 (**1), 77-88.

_________ (1911b) Review of *An Introduction to the Theory of Statistics* by G. Udny Yule, *Economic Journal*, **21** (Jun.), 263-264.

_________ (1913) Working-Class Households in Reading, *Journal of the Royal Statistical Society*, **76 (**7), 672-701.

_________ (1915) Criticism of the Accuracy of the Results, chapter VI and pp.

174-184 of Bowley & Burnett-Hurst (1915).

_________ (1920) *Elements of Statistics,* 4th edition (Expanded and substantially revised), London: King.

_________ (1922) Studies in Probability. II. The Logic of Probability and Statistics, *Economica*, No. 4. 97-100.

_________ (1923) The Precision of Measurements based on Samples, *Metron*, **2** (3), 494-500.

_________ (1924) Contribution to the discussion of Hilton (1924), *Journal of the Royal Statistical Society*, **87** (4), 563-565.

_________ (1925) Note on the Method of Sampling, pp. 48-51 of Bowley & Hogg (1925).

_________ (1926a) Contribution to the discussion of Jensen (1926a), *Bulletin de l'Institut International de Statistique*, **22**, Suppl. to Book 1, 62-64.

_________ (1926b) Measurement of the Precision Attained in Sampling, *Bulletin de l'Institut International de Statistique*, **22**, Suppl. to Book 1, 1-62.

_________ (1926c) The Influence on the Precision of Index-Numbers of Correlation Between the Prices of Commodities, *Journal of the Royal Statistical Society*, **89** (2), 300-319.

_________ (1928a) *F. Y. Edgeworth's Contribution to Mathematical Statistics*, Royal Statistical Society, London. Reprinted 1972 by Kelley, Clifton, NJ.

_________ (1928b) Notes on Index Numbers, *Economic Journal*, **38**, 216-237.

_________ (1932a) The House Sample Analysis. Pages 29-96 of Smith (ed.) *The New Survey of London Life and Labour, volume 3: Survey of Social Conditions: 1. The Eastern Area.*

_________ (1932b) Appendix IV: The House Sample Analysis. Pages 439-448 of Smith (ed.) *The New Survey of London Life and Labour, volume 3: Survey of Social Conditions: 1. The Eastern Area.*

_________ (1934a) Francis Ysidro Edgeworth, *Econometrica*, **2** (2), 113-124.

_________ (1934b) Contribution to the discussion of Neyman (1934), *Journal of the Royal Statistical Society*, **97** (4), 607-610.

_________ (1934c) The House Sample Analysis. Pages 29-117 of Smith (ed.) *The New Survey of London Life and Labour, volume 6: Survey of Social Conditions: 2. The Western Area.*

_________ (1935) Contribution to the discussion of Fisher (1935), *Journal of the Royal Statistical Society*, **98** (1), 55-57.

_________ (1936) The Application of Sampling to Economic and Sociological Problems, *Journal of the American Statistical Association*, **31** (September), 474-480.

_________ (1937) *Elements of Statistics,* 6th edition (With a supplement added to the text of the 5th edition), London: King.

_________ (1952) Index-Numbers of Wage-Rates and Cost of Living, *Journal of the Royal Statistical Society A*, **115 (**4), 500-506.

_________ (1955) B. Seebohm Rowntree, C. H. 1871-1954, *Journal of the Royal Statistical Society A*, **118 (**1), 125-126.

Bowley, A. L. & A. R. Burnett-Hurst (1915) *Livelihood and Poverty: A Study in the Economic Conditions of Working-class Households in Northampton, Warrington, Stanley and Reading*, London: Bell.

Bowley, A. L. & L. R. Connor (1923) Tests of Correspondence between Statistical Grouping and Formulae, *Economica*, **(7)**, 1-9.

Bowley, A. L. & F. Y. Edgeworth (1902) Methods of Representing Statistics of Wages and Other Groups not Fulfilling the Normal Law of Error, *Journal of the Royal Statistical Society*, **65 (**2), 325-354.

Bowley, A. L. & M. H. Hogg (1925) *Has Poverty Diminished?* London: King.

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

Bulmer, M., K. Bales and K. K. Sklar (eds.) (1991) *The Social Survey in Historical Perspective, 1880-1940*. Cambridge: Cambridge University Press.

Chang, W.-C. (1976) Sampling Theories and Practice. In D. B. Owen (ed.) *On the History of Statistics and Probability*, pp. 299-315. New York: Wiley.

Chuprov, A. A. (Tchouproff) (1918) On the Mathematical Expectation of the Moments of Frequency Distributions, *Biometrika*, **12** (1/2), 140-169.

_________ (1919) On the Mathematical Expectation of the Moments of Frequency Distributions, *Biometrika*, **12** (3/4), 185-210

_________ (1921) On the Mathematical Expectation of the Moments of Frequency Distributions. Part II, *Biometrika*, **13** (2/3), 283-295.

_________ (Tschuprow) (1923) On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations, *Metron*, **2** (3), (4), 461-493, 646-683.

_________ (Tschuprow) (1925) On the Asymptotic Frequency Distribution of the Arithmetic Means of n Correlated Observations for Very Great Values of n, *Journal of the Royal Statistical Society*, **88** (1), 91-104.

_________ (1926) Contribution to the discussion of Jensen (1926), *Bulletin de l'Institut International de Statistique*, **22**, Suppl. to Book 1, 1-62.

Clapham, A. R. (1929) The Estimation of Yield in Cereal Crops by Sampling Methods, *Journal of Agricultural Science*, **19**, 214-235.

Clopper, C. J. and E. S. Pearson (1934) The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, *Biometrika*, **26** (4), 404-413.

Dale, A. I. (1999) *A History of Inverse Probability from Thomas Bayes to Karl Pearson*, second edition. New York: Springer-Verlag.

_________ (2001) Arthur Lyon Bowley. In C. C. Heyde and E. Seneta (eds.) *Statisticians of the Centuries*, pp. 278-282. New York: Springer.

Darnell, A. C. (1981) A. L. Bowley 1869-1957, pp. 140-174 of D. P. O'Brien and J. R. Presley (eds.) *Pioneers of Modern Economics in Britain.* London: Macmillan.

Desrosières, A. (1991) The Part in Relation to the Whole–How to Generalize?; the Prehistory of Representativeness. In M. Bulmer, K. Bales, K. K. Sklar (eds.) (1991), pp. 140-174.

Dwyer, P. S. (1938) On Combined Expansions of Products of Symmetric Power Sums and of Sums of Symmetric Power Products with Applications to Sampling, (in 2 parts), *Annals of Mathematical Statistics*, **9** (1), (2), 1-47, 97-132.

Edwards, A. W. F. (1997) What Did Fisher Mean by "Inverse Probability" in 1912-1922? *Statistical Science*, **12** (3), 177-184.

Edgeworth, F. Y. (1884a) Philosophy of Chance, *Mind*, **9** (), 223-235.

_________ (1884b) À Priori Probabilities, *Philosophical Magazine*, **17** (), 204-210.

_________ (1885) Methods of Statistics, *Jubilee Volume, Royal Statistical Society*, pp. 181-217.

_________ (1892) The Law of Error and Correlated Averages, *Philosophical Magazine*, **34**, 429-438.

_________ (1906) The Generalised Law of Error, or Law of Great Numbers, *Journal of the Royal Statistical Society*, **69** (3), 497-539.

_________ (1908-9) On the Probable Errors of Frequency-Constants, *Journal of the Royal Statistical Society*, **71** (2) (3) (4), 381-397, 499-512, 651-678; **72** (1), 81-90.

_________ (1913) On the Use of the Theory of Probabilities in Statistics Relating to Society, *Journal of the Royal Statistical Society*, **76** (2), 165-193.

_________ (1921) Molecular Statistics, *Journal of the Royal Statistical Society*, **84** (1), 71-89.

_________ (1924) Contribution to the discussion of Hilton (1924), *Journal of the Royal Statistical Society*, **87** (4), 565-568.

Edgeworth, F. Y. and A. L. Bowley (1902) Methods of Representing Statistics of Wages and Other Groups not Fulfilling the Normal Law of Error, *Journal of the Royal Statistical Society*, **65** (2), 325-354.

Ericson, W. A. (1969) Subjective Bayesian Models in Sampling Finite Populations, *Journal of the Royal Statistical Society B*, **31** (2), 195-233.

Fienberg, S. E. (2006) When did Bayesian Inference become "Bayesian"? *Bayesian Analysis*, **1** (1), 1-40.

Fienberg, S. E. and Tanur, J. M. (1996) Reconsidering the Fundamental Contributions of Fisher and Neyman on Experimentation and Sampling, *International Statistical Review*, **64** (3), 237-253.

_________ (2001) History of Sample Surveys. In N. J. Smelser and P. B. Baltes (ed.) *International Encyclopedia of Social and Behavioral Science*, vol. ?, pp. 13453-13458. Amsterdam: Elsevier.

Fisher, R. A. (1922) On the Mathematical Foundations of Theoretical Statistics, *Philosophical Transactions of the Royal Society, A*, **222**, 309-368.

_________ (1922a) On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P, *Journal of the Royal Statistical Society*, **85**, 87-94.

_________ (1923) Statistical Tests of Agreement between Observation and Hypothesis, (with a comment by Bowley 146-7), *Economica*, (8), 139-147.

_________ (1924) The Conditions under which χ^2 Measures the Discrepancy between Observation and Hypothesis, *Journal of the Royal Statistical Society*, **87 (**?), 442-450.

_________ (1925a) *Statistical Methods for Research Workers*, Edinburgh: Oliver & Boyd.

_________ (1925b) Theory of Statistical Estimation, *Proceedings of the Cambridge Philosophical Society*, **22**, 700-725.

_________ (1930) Inverse Probability, *Proceedings of the Cambridge Philosophical Society*, **26**, 528-535.

_________ (1934) Contribution to the discussion of Neyman (1934), *Journal of the Royal Statistical Society*, **97 (**4), 614-619.

_________ (1935) The Logic of Inductive Inference, (with discussion) *Journal of the Royal Statistical Society,* **98 (**1), 39-82.

Gini, C. and Galvani, (1929). Di una Applicazione del Metoda Rappresentative all'utimo Censimento Italiano della Populazione, *Annali di Statistica*, **4**, 1-107.

Greenwood, M. and L. Isserlis (1927) A Historical Note on the Problem of Small Samples, *Journal of the Royal Statistical Society*, **90** (2), 347-352.

Groenewegen, P. (1995) *A Soaring Eagle: Alfred Marshall 1942-1924*, Cheltenham, Glos.: Elgar.

Hald, A. (1998) *A History of Mathematical Statistics from 1750 to 1930*, New York: Wiley.

_________ (1999) On the History of Maximum Likelihood in Relation to Inverse Probability and Least Squares, *Statistical Science*, **14** (2), 214-222.

Hansen, M. H., Dalenius, T., and Tepping, B. J. (1985) The Development of Sample Surveys of Finite Populations. In A. C. Atkinson and S. E. Fienberg (eds.) *A Celebration of Statistics. The ISI Centenary Volume*, pp. 327–354. New York: Springer.

Hansen, M. H. and W. G. Madow (1976) Some Important Events in the Historical Development of Sample Surveys, In D. Owen (ed.) *On the History of Statistics and Probability*, New York: M. Dekker.

Hennock, E. P. (1987) The Measurement of Urban Poverty: from the Metropolis to the Nation, 1880–1920, *Economic History Review*, 2nd ser., **40**, 208–27.

Hepple, L. W. (2001) Multiple Regression and Spatial Policy Analysis: George Udny Yule and the Origins of Statistical Social Science, *Environment and Planning*

D: Society and Space, **19**, 385-407.

Hilton, J. (1924) Enquiry by Sample: An Experiment and Its Results, *Journal of the Royal Statistical Society*, **87** (4), 544-570.

Hubback, J. A. (1927) Sampling for Rice Yield in Bihar and Orissa, *Bulletin No 166*, Imperial Agricultural Research Institute, Pusa, India. Reprinted in *Sankhyā*, **7**, (1946), 281-294.

Irwin, J. O. (1966) Leon Isserlis, M.A., D.Sc. (1881-1966), *Journal of the Royal Statistical Society A*, **129** (4), 612-616.

Isserlis, L. (1918) On the Value of a Mean as Calculated from a Sample, *Journal of the Royal Statistical Society*, **81** (1), 75-81.

__________ (1926) Alexander Alexandrovitch Tschuprow, *Journal of the Royal Statistical Society*, **89** (3), 619-622.

Jeffreys, H. (1939) *Theory of Probability*, Oxford, University Press.

Jensen, A. (1923) Méthodes permettant de réaliser une économie de travail dans la statistique, presented at the ISI Session in Brussels 1923, Copenhagen: Bianco Luno.

__________ (1926a) Report on the Representative Method in Statistics, *Bulletin de l'Institut International de Statistique*, **22**, Book 1, 359–380.

__________ (1926b) The Representative Method in Practice, *Bulletin de l'Institut International de Statistique*, **22**, Book 1, 381–439.

_________ (1928) Purposive Selection, *Journal of the Royal Statistical Society*, **91** (4), 541-547.

Keynes, J. M.(1909) The Method of Index Numbers with Special Reference to the Measurement of General Exchange Value, Adam Smith Prize Essay, first published in *The Collected Writings of John Maynard Keynes, XI* 1983 London: Macmillan for the Royal Economic Society.

_________ (1921) *A Treatise on Probability*, London, Macmillan.

_________ (1930) *A Treatise on Money in Two Volumes: 1 The Pure Theory of Money*, London: Macmillan.

Kotz, S. and A. I. Dale (2009) *Arthur L. Bowley: A Pioneer in Modern Statistics and Economics*, Singapore: World Publishing.

Kruskal, W. and Mosteller, F. (1980) Representative Sampling. IV. The History of the Concept in Statistics, 1895-1939, *International Statistical Review*, **48**, 169-195.

Lexis, W. (1877) *Theorie der Massenerscheinungen in der menschlichen Gesellschaft*, Freiburg.

Mahalanobis, P. C. (1944) On Large-Scale Sample Surveys, *Philosophical Transactions of the Royal Society B*, **231**, (Oct. 31), 329-451.

_________ (1946) Sample Surveys of Crop Yields in India, *Sankhyā*, **7**, 269-280.

March, L. (1926) Observations sur la Méthode Représentative et sur le Projet de Rapport Relatif à cette Méthode, *Bulletin de l'Institut International de Statistique*,

22, Liv. 1, 444-451.

Maunder, W. F. (1972) Sir Arthur Lyon Bowley (1869-1957), an inaugural lecture delivered at the University of Exeter. Reprinted in M. G. Kendall and R. L. Plackett (eds.) *Studies in the History of Statistics and Probability II*, 1977. London: Griffin.

Merriman, M. (1899) *A Text-book on the Method of Least Squares*, seventh edition (first edition 1884), New York: Wiley.

Mises, R. von (1919) Grundlagen der Wahrscheinlichkeitsrechnung, *Mathematische Zeitschrift*, **5**, 52-99.

Moser, C. A. (1949) The Use of Sampling in Great Britain, *Journal of the American Statistical Association*, **44** (June), 231-259.

Neyman, J. (Splawa-Neyman) (1925) Contributions to the Theory of Small Samples Drawn from a Finite Population, *Biometrika*, **17**, (3/4), 472-479.

_________ (1929) Contributions to the Theory of Certain Test Criteria, *Bulletin de l'Institut International de Statistique*, **49**, 3-48.

_________ (1934) On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection, (with discussion), *Journal of the Royal Statistical Society*, **97** (4), 558-625.

Neyman, J. and E. S. Pearson (1928) On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference: Parts I and II, *Biometrika*,

20A, (1/2 and 3/4), 175-240 and 263-294.

_________ (1933) The Testing of Statistical Hypotheses in Relation to Probabilities A Priori, *Proceedings of the Cambridge Philosophical Society*, **29**, 492-510.

Pearson, K. (1896) Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity and Panmixia, *Philosophical Transactions of the Royal Society A*, **187**, 253-318.

_________ (1900) On the Criterion that a Given System of Deviations from the Probable in the Case of Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling, *Philosophical Magazine*, **50**, 157-175.

_________ (1903) Editorial: On the Probable Errors of Frequency Constants, *Biometrika*, **2** (3), 273-281.

_________ (1907) On the Influence of Past Experience on Future Expectation, *Philosophical Magazine*, **13**, 365-378.

_________ (1914) *Tables for Statisticians and Biometricians*, Cambridge: Cambridge University Press.

_________ (1920) The Fundamental Problem of Practical Statistics, *Biometrika,* **13** (1), 1-16.

_________ (1928) On a Method of Ascertaining Limits to the Actual Number of Marked Members in a Population of Given Size from a Sample, *Biometrika,* **20A**

(1/2), 149-174.

_________ (1929) Laplace, *Biometrika*, **21**(1/4), 202-216.

Porter, T. M. (1986) *The Rise of Statistical Thinking 1820-1900*, Princeton: Princeton University Press.

Pratt, J. W. (1976) F. Y. Edgeworth and R. A. Fisher on the Efficiency of Maximum Likelihood Estimation, *Annals of Statistics*, **4**, 501-514.

Rao, J. N. K. (2005) Interplay Between Sample Survey Theory and Practice: An Appraisal, *Survey Methodology* , **31**, (2), 117-138.

Reid, C. (1982) *Neyman–from Life*, New York: Springer.

Rowntree, B. S. (1901) *Poverty, A Study of Town Life*, London: Macmillan.

Seneta, E. (1985) A Sketch of the History of Survey Sampling in Russia, *Journal of the Royal Statistical Society*, **148 (**2), 118-125.

Seng, Y. P. (1951) Historical Survey of the Development of Sampling Theories and Practice, *Journal of the Royal Statistical Society A*, **114 (**2), 214-231.

Smith, H. L. (1930-35) *The New Survey of London Life and Labour*, 9 viols. London: King.

Smith, T. M. F. (1976) The Foundations of Survey Sampling: A Review (with discussion), *Journal of the Royal Statistical Society A*, **139 (**2), 183-204.

Stephan, F. F. (1948) History of the Uses of Modern Sampling Procedures, *Journal of the American Statistical Association*, **43**, 12-39.

Stigler, S. M. (1978) Francis Ysidro Edgeworth, Statistician, (with discussion) *Journal of the Royal Statistical Society*, **141** (3), 287-322.

_________ (1986) *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA, and London: Belknap Press of Harvard University Press.

_________ (2005) Fisher in 1921, *Statistical Science*, **20** (1), 32-49.

'Student' (1908) Probable Error of a Correlation Coefficient, *Biometrika*, **6**, 302-310.

Sugden, R. A., T. M. F. Smith and R. P. Jones (2000) Cochran's Rule for Simple Random Sampling, *Journal of the Royal Statistical Society B*, **62** (4), 787-793.

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*, first and second editions in 1866 and -76, London: Macmillan.

Wrinch, D. & H. Jeffreys (1919) On Some Aspects of the Theory of Probability, *Philosophical Magazine*, **38**, 715-731.

Yates, F. (1946) A Review of Recent Statistical Developments in Sampling and Sampling Surveys, *Journal of the Royal Statistical Society*, **109** (1), 12-43.

Yule, G. U. (1896) Notes on the History of Pauperism in England and Wales from 1850, Treated by the Method of Frequency-Curves; with an Introduction on the

Method, *Journal of the Royal Statistical Society*, **59** (2), 318-357.

__________ (1911) *An Introduction to the Theory of Statistics*, (third edition in 1922), London: Griffin.

__________ (1921) Review of *Elements of Statistics* by A. L. Bowley, *Economic Journal*, **31**, No. 122 (Jun.), 220-224.

__________ (1924) Contribution to the discussion of Hilton (1924), *Journal of the Royal Statistical Society*, **87** (4), 562-563.

Zabell, S. (1989) R. A. Fisher on the History of Inverse Probability (with discussion), *Statistical Science*, **4** (3), 247-263.

__________ (1992) R. A. Fisher and the Fiducial Argument, *Statistical Science*, **7** (3), 369-387.