International Statistical Review (3), 1966, 64, 237–253, Printed in Mexico

Reconsidering the Fundamental Contributions of Fisher and Neyman on Experimentation and Sampling

Stephen E. Fienberg[1] and Judith M. Tanur[2]

[1]*Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*
[2]*Department of Sociology, State University of New York at Stony Brook, Stony Brook, NY 11794-4356, USA*

Summary

R.A. Fisher and Jerzy Neyman are commonly acknowledged as the statisticians who provided the basic ideas that underpin the design of experiments and the design of sample surveys, respectively. In this paper, we reconsider the key contributions of these great men to the two areas of research. We also explain how the controversy surrounding Neyman's 1935 paper on agricultural experimentation in effect led to a split in research on experiments and on sample surveys.

Key words: Clustering; Design of experiments; Design of sample surveys; Randomization; Stratification.

1 Introduction

The 1920s and 1930s mark a critical period in the development of statistics. Much of the modern theory of estimation and statistical testing emerged from papers published during this period by two of the most important statisticians of the century, Ronald Aylmer Fisher and Jerzy Neyman. In many ways, the theoretical work of these two great men was stimulated by practical problems they encountered in their day-to-day statistical work of agricultural experimentation and, later for Neyman, the sampling of human populations. Moreover many believe that their greatest accomplishments were not in the realm of theories of inference but rather in their articulation of theory and principles underpinning the two key modern methods for the scientific collection of data—randomized experiments and random sampling.

In connection with a larger project on experiments and sample surveys, and in part stimulated by the recent occasion of the 100th anniversary of Neyman's birth, we have reread much of the work of both Fisher and Neyman in the areas of sampling and experimentation. Our story for both Fisher and Neyman begins with the work that led to their 1923 papers on agricultural experimentation. For Neyman, this work continued largely in the direction of sampling and culminated in his now famous 1934 paper on the topic. Despite the many contributions of others to the technical development of sampling in the period from 1900 to 1925, Neyman's 1934 paper played a pivotal role in turning the dry mathematics of expectations into real sampling plans for actual randomly selected large scale surveys. For Fisher, his 1923 paper led to further pioneering work in experimental design that culminated in the publication of his 1935 book, *The Design of Experiments*. Fisher's book played a catalytic role in the actual use of randomization in controlled experiments that is similar to the role Neyman's 1934 paper played in the use of methods for random sampling. The year 1935 was a critical one in the developments in the fields of experimentation and sampling, not simply because of the publication of Fisher's book but also because of a major controversy between Fisher and Neyman engendered by Neyman's (1935) paper on agricultural experimentation. Because of the bitterness that grew out of this dispute (and a related one between Fisher on the one hand and Neyman and Pearson on the other, over tests of hypotheses and then later over confidence intervals), Fisher and Neyman were never able to bring their ideas together and benefit from the fruitful interaction that would likely have occurred had they done so. And in the aftermath, Neyman staked out intellectual responsibility for sampling while Fisher did the same for experimentation. It was in part because of this rift between Fisher and Neyman that the fields of sample surveys and experimentation drifted apart.

In the next section, we begin with Neyman's important 1923 paper on agricultural experimentation, and we remind the reader of the interrelated nature of fundamental ideas on experiments and surveys and the pivotal role that randomization played in both areas. Then our narrative proceeds by interweaving accounts of the work by Fisher and Neyman on experimentation and sampling, and controversies that surrounded them. Our account culminates with the clash between Fisher and Neyman over ideas in Neyman's 1935 paper. Throughout we include relevant biographical material assembled from a variety of sources including Box (1978, 1980), Lehmann & Reid (1982) and Reid (1982).

2 Parallels between Surveys and Experiments: Innovations in Neyman's 1923 Work on Agricultural Experimentation

Jerzy Neyman was born of Polish parents in 1894, in Bendery, which has been variously labeled as Rumania, Ukraine, and Moldavia because of the vicissitudes of border-drawing in Eastern Europe. In 1912 he entered the University of Kharkov (which later became Maxim Gorki University) to study mathematics. Finishing his undergraduate studies in 1917, Neyman remained at the University of Kharkov to begin to prepare for an academic career; he also received an appointment as a lecturer at the Kharkov Institute of Technology. In the fall of 1920 he passed the examination for a Masters degree and became a lecturer at the University. Thus until after his 27th birthday Neyman was doubly isolated—both by living in a provincial city and by being part of an ethnic minority in that city. In the spring of 1921, learning that he was to be arrested, Neyman fled to the country home of a relative, where he supported himself by teaching the children of peasants. In the summer he returned to Kharkov and thence to Bydgoszcz in northern Poland, as part of an exchange of nationals between Russia and Poland agreed to at the end of the Russian–Polish War. In Bydgoszcz, Neyman went to work as "senior statistical assistant" at the National Agricultural Institute and it was there that he wrote two long papers on agricultural experimentation that were published in 1923 in Polish (Splawa–Neyman, 1990 [1923a], 1925 [1923b]). In this section we focus primarily on the first of these papers, and we return to the second in Section 4, discussing its 1925 republication in English.

An excerpt of the 1923 paper (Splawa–Neyman, 1990 [1923a]) on experimentation was recently translated from the Polish original and published in *Statistical Science*, and in it we were especially struck by the importance that repeated random sampling played in Neyman's thinking. This seemed to us to foreshadow, at the very least, the use of randomization in experimentation. Reid (1982, p. 44) quotes Neyman considerably later as denying his priority here:

> "... I treated theoretically an unrestrictedly randomized agricultural experiment and the randomization was considered as a prerequisite to probabilistic treatment of the results. This is not the same as the recognition that without randomization an experiment has little value irrespective of the subsequent treatment. The latter point is due to Fisher and I consider it as one of the most valuable of Fisher's achievements."

Since we see one of the major purposes of experimental randomization as the necessary precondition for probabilistic inference from the results, we would join Rubin (1990, p. 477) in saying that had Neyman later claimed priority rather than denying it, we would have had no reason to quarrel with that claim. Rubin (1990) also reminds us that the use of randomization was "in the air" in the early 1920s, citing Student (1923, pp. 281–282) and Fisher & MacKenzie (1923, p. 473).

There is, however, another important feature of Neyman's first paper on the topic of agricultural experimentation we think especially worthy of note. For a number of years, we have pursued the parallels and linkages between surveys and experiments (e.g., see Fienberg & Tanur, 1987, 1988, 1989) and, in particular, Fisher's and Neyman's roles as progenitors of ideas in both areas. Although in fact surveys and experiments had developed very long and independent traditions by the start of the 20th century, it was only with the rise of ideas associated with mathematical statistics in the 1920s that the tools for major progress in these areas became available. The key intellectual idea was the role of randomization or random selection, both in experimentation and in sampling, and both Neyman and Fisher utilized this idea, although in different ways. But then, as now, it is clear that one can adapt the ideas from one field to the other and that one can profitably link them, as in sampling embedded within experiments or experiments embedded within surveys.

Neyman's first 1923 paper on experimental design embodies this sort of adaptation. He conceptualizes the assignment of treatments to units in an experiment as the drawing without replacement of balls from urns, one urn for each treatment. These urns have the special property that the removal of a ball (representing the outcome of an experimental unit) from one urn causes it to disappear from the other urns as well. Thus Neyman shows that when there is a finite pool of experimental units that need to be assigned to treatments, the random assignment of units to treatments is exactly parallel to the random selection of a sample from a finite population. Hence, when the number of units used in an experiment is a large fraction of the units in the population, a finite population correction must be used in an experiment, just as it is in a sample survey. The parallel between experiments and sampling is particularly close in this case. Nevertheless, all but a few modern investigators have lost sight of the parallel and fail to take advantage of insights offered in the parallel literature.

3 Fisher's Initial Contributions to the Design of Experiments in the 1920s

While it is only recently that statisticians and others have rediscovered Neyman's early contributions to the design of experiments, Fisher has long been recognized for the innovation he brought to the field.

Fisher was born in 1890 in London and it was during his undergraduate years at Cambridge that his interests in eugenics, genetics, and biometry stimulated his interest in probability and statistics. After graduation in 1913, he taught mathematics and physics at a series of schools but pursued research in both genetics and statistics and published his first major papers on these topics. This work led, six years later, to a temporary statistical position in 1919 at Rothamsted Experiment Station, where the director, Sir John Russell, wanted someone who would be prepared to examine the accumulated data on wheat yields from Broadbalk Fields in order to elicit further information that might have been missed. Russell quickly recognized Fisher's genius and set about converting the temporary position to a more permanent one (see Box, 1978, pp. 96–97). In 1921, Fisher published the first of a series of papers entitled "Studies in crop variation" intended for the *Journal of Agricultural Science*[1], in which he presented some of his reanalyses of the Broadbalk data. But it was in the second in this series of papers, written with W.A. Mackenzie (1923), that Fisher's new ideas on agricultural experimentation emerged.

Fisher & Mackenzie (1923) contains two key innovations for the design of experiments: the

[1]Actually the third paper in the series appeared not in the *Journal of Agricultural Science* but in the *Philosophical Transactions of the Royal Society*, and without the title, "Studies in crop variation". Two later papers with Thomas Eden in 1927 and 1929 are labelled as V and VI.

introduction of the analysis of variance, adapted from Fisher's 1918 paper on Mendelian inheritance, and randomization. In fact, in introducing the analysis of variance, the authors argue that it is conditional on randomization:

> "Further, if all the plots are undifferentiated, as if the numbers had been mixed up and written down in random order, the average value of each of the two parts is proportional to the number of degrees of freedom in the variation of which it is compared." (p. 315)

The spirit of this statement is not unlike that in Splawa–Neyman (1990 [1923a]): "The goal of a field experiment which consists of the comparison of n varieties will be regarded as equivalent to the problem of comparing the numbers $a_1, a_2, \dots, a_n$ or their estimates by way of drawing several balls from an urn." (p. 467) Both speak of randomization, but in hypothetical terms. There are dramatic differences between Fisher's and Neyman's treatments as well, with Fisher focusing on the actual data analysis and Neyman putting much emphasis on the mathematical formulation of finite sampling and expectations.

Neither Fisher nor Neyman actually wrote about the implementation of randomization in 1923, however. In fact, as Cochran (1980, p. 18) notes in connection with the $2 \times 12 \times 3$ factorial experiment described by Fisher & Mackenzie (1923):

> "No randomization was used in this layout. Following the procedure recommended at that time, the layout apparently attempts to minimize the errors of the differences between treatment means by using a chessboard arrangement that places different treatments near one another as far as is feasible. This arrangement utilizes the discovery from uniformity trials that plots near one another in a field tend to give closely similar yields. A consequence is, of course, that the analysis of variance estimate of the error variance per plot, which is derived from differences in yield per plots receiving the same treatment, will tend to overestimate, since plots treated alike are farther apart than plots receiving different treatments. Fisher does not comment on the absence of randomization or on the chessboard design. Apparently in 1923 he had not begun to think about the conditions necessary for an experiment to supply an unbiased estimate of error."

A further problem with Fisher's 1923 analysis was his failure to recognize the split-plot structure of the actual experiment with respect to a third factor, potassium. This he corrected in *Statistical Methods for Research Workers* (1925).

Thus, in 1923, we see both Fisher and Neyman as having arrived at similar positions with respect to randomization, although Fisher already had moved ahead in his thinking about other features of design. Box (1980) argues, that the statement in Fisher & Mackenzie on the analysis of variance rested heavily on Fisher's appreciation of the requirements of underlying theory for the normal distribution upon which he relied in his analysis, as well as upon Fisher's "geometric representation that by then was second nature to him. He could picture the distribution of n results as a pattern in n-dimensional space, and he could see that randomization would produce a symmetry in that pattern rather like that produced by a kaleidoscope, and which approximated the required spherical symmetry available, in particular, from standard normal theory assumptions" (p. 2). While we agree that Fisher's geometrical intuition was highly refined, we see little support in the 1923 paper for this conclusion.

But by the publication of *Statistical Methods* in 1925, Fisher clearly had a deeper appreciation of the function of randomization in experimentation:

> "The first requirement which governs all well-planned experiments is that the experiment should yield not only a comparison of different manures, treatments, varieties, etc., but also a means of testing the significance of such differences as are observed .... For our test of significance to be valid the differences in fertility between plots chosen as

> parallels must be truly representative of the differences between plots with different treatment[s]; and we cannot assume that this is the case if our plots have been chosen in any way according to a prearranged system." (p. 248)

According to Fisher, valid tests result if treatments are assigned to plots wholly at random, and here we have a clear departure from the 1923 papers.

The treatment of experimentation in *Statistical Methods*, which follows this discussion of randomization, consists of only a few pages. Fisher introduced blocking as a device to increase the accuracy of treatment comparisons and he illustrated its use in both randomized block experiments and in Latin squares. He described the split-plot structure of the experiment analyzed in the 1923 paper, which should have required two levels of randomization for its validity, but Fisher did not discuss this point in the original or subsequent editions of the book.

Later embellishments to the theory of experimental design came in Fisher (1926) where he laid out in a more systematic fashion the principles underlying field experimentation, including the need for replication, and in which he expounded upon the notion of factorial designs with new insights, including the role of confounding. The example used in this paper was from a real experiment being run by a colleague, Thomas Eden, and Eden & Fisher (1927), presents a more detailed description of the experiment and an analysis of its results. Factorial structures, Fisher (1952) would later observe, had their roots in the "simultaneous inheritance of Mendelian factors".

Throughout this period, Fisher continued to work with colleagues at Rothamsted and elsewhere in the design and analysis of experiments that implemented his ideas on randomization and factorial structures. But not all of his Rothamsted colleagues had fully accepted Fisher's insistence on randomization as the foundation of experimentation (Box, 1980). In fact, Fisher's 1926 article was at least in part a response to an article by Sir John Russell, director of Rothamsted, on the state of the art in experimental design. In his paper, Russell had noted Fisher's argument for randomization but said that it was impossible to use in practice! Despite this disagreement, Russell allowed Fisher and Eden to run their complex randomized designs on fields at Rothamsted and this ultimately convinced many others of the value of Fisher's innovative ideas (Box, 1980).

4 Neyman's Initial Contributions to the Theory of Sampling in the 1920s

In 1924, Neyman obtained his doctors degree from the University of Warsaw, using as a thesis the work done in Bydgoszcz, and in 1925, he obtained a government grant to study for a year in London with Karl Pearson. Before his arrival in London, Neyman had shipped to Karl Pearson several of his statistical publications, including the two 1923 papers on agricultural experimentation, and Pearson suggested that Neyman republish part of the second paper in English in *Biometrika* (Splawa–Neyman, 1925 [1923b]). But Pearson believed that Neyman's statement at the end of the paper, that it is only in sampling from a normal population that the sample mean and sample variance are independent, to be mistaken. When Neyman tried to explain to Pearson his confusion between independence and lack of correlation (in halting English and in front of several other Pearson students), Pearson interrupted: "That may be true in Poland, Mr. Neyman, but it is not true here." (Reid, 1982, p. 57) Dismayed at having offended Pearson and at perhaps having lost a chance to publish in English—his Polish mentors had sent Neyman to England as a kind of test to see if his ideas were worth anything, and a publication in English would go a long way towards settling that matter—Neyman searched for a way to communicate his explanation to Pearson. He finally offered his explanation to J.O. Irwin, who communicated it to Egon Pearson, who finally convinced his father that Neyman was not mistaken. Thus the *Biometrika* version does contain this observation from the Polish version.

In the resulting 1925 *Biometrika* paper, Neyman gave the higher moments of the means and variances of samples from finite populations. The paper is succinct and to the point. He begins by deriving formulas for moments of the sampling distribution of the mean of samples of size n from a finite population of size m, where the sample members are drawn without replacement, and he relates these to the moments of the underlying finite population. Then he focuses on the variance, skewness, and kurtosis, of the sampling distribution, i.e., M_2, $B_1 = M_3^2/M_2^3$, and $B_2 = M_4/M_2^2$, where M_2, M_3, and M_4 are the central moments of the sampling distribution, and he expresses these in terms of μ_2, $\beta_1 = \mu_3^2/\mu_2^3$, and $\beta_2 = \mu_4/\mu_2^2$ where μ_2, μ_3, and μ_4 are the central moments of the finite population of m elements. He explains how the moments of the sampling distribution tend to the usual ones for sampling with replacement when we let $m \to \infty$. He then turns to the first and second moments of the sample variance, expressing them in terms μ_2 and β_2, and to the correlation between the square of the sample mean and the sample variance as well as to the correlation between the sample mean and the sample variance, expressing the first in terms of β_2 and the second in terms of β_1 and β_2. In the final paragraph of the paper, by letting the finite population size $m \to \infty$, Neyman uses the correlation formulas to argue that the independence of the sample mean and the sample variance holds only for normal distributions, the point originally in dispute with Pearson.

The 1925 paper might have passed unnoticed, but for the fact that, in 1927, Major Greenwood and Leon Isserlis published a complaint about it in the *Journal of the Royal Statistical Society*, pointing out that Neyman had failed to acknowledge the published papers of the recently deceased Russian statistician, Alexander Alexandrovitch Tchouproff[2].

Greenwood & Isserlis (1927, p. 348) quote Neyman as writing of the formula for the second moment of the variance "[t]his result, being a generalization of formulae given by other authors, is, I believe, novel and of considerable importance". Indeed, after that statement on page 477 of his 1925 paper, Neyman did actually cite one of Tchouproff's 1918 *Biometrika* papers as well as a 1925 paper by Church. Greenwood and Isserlis nevertheless take him to task for failing to cite works by Karl Pearson, Isserlis, and Edgeworth besides those of Tchouproff. In their view, Neyman's felony was compounded by the work of A.E.R. Church, especially a 1926 paper published in *Biometrika*, which cited Neyman for some of the formulas for moments rather than citing Tchouproff. There is little question but that Neyman's results can be found at least in some form in the sea of formulas in Tchouproff's 1923 papers. What appears to have been happening in these and other papers by Tchouproff is the alternative algebraic derivation and expression of a variety of moment expressions, most of which were quite complex. One of the nice features of Neyman (1925) is the relatively clean derivations and succinct formulas.

By the time the Greenwood–Isserlis critique was published, Neyman had returned to Poland. In obvious distress at the accusation, he wrote to Egon Pearson soliciting advice. The reply, it turned out, was handled by Karl Pearson (1927), who editorialized in *Biometrika* in Neyman's defense (as well as Church's). He argued that Neyman's original publication in Polish in 1923 was certainly contemporaneous with Tchouproff's pair of 1923 articles in *Metron*, and perhaps actually predated those Tchouproff publications because delays in the publication of *Metron* caused it to appear later than its cover date. But it seems to us that this argument from Pearson, while it may well be true, is irrelevant to the controversy—Greenwood and Isserlis were accusing Neyman of ignoring not these specific papers of Tchouproff but a whole body of literature that originated some 20 years earlier with work by Pearson himself. Perhaps Pearson was in some sense defending himself as editor of *Biometrika* for his laxness in failing to urge Neyman, the young foreigner, to update his work for an English-reading audience with citations to the literature in English. That Neyman could have profited by such urging seems clear from another facet of Pearson's editorial defense—he cites a letter from Dr. K. Bessalik, Professor of the University of Warsaw, who in 1922 was Director of the State Institute of Agricultural Research in Bydgozcz. In that letter, Bessalik certifies that Neyman wrote his paper in Bydgozcz in 1922 and that "no English Journals" were accessible to him. Had Pearson urged Neyman to place the republication of his results in the historical context about which he had

[2]This was the transliteration of Tchouproff's name, as it appeared in the 1918 *Biometrika* articles. We use the spelling Tchouproff throughout, even though when he published in *Metron*, the editor used Tschuprow, and elsewhere he is referred to as Chuprov and Chouprow, but we refer to specific papers using the spelling used in connection with them.

presumably been ignorant during the time he was writing the original paper in Poland, we believe that Neyman would undoubtedly have agreed. He had, after all, come to London with the expressed purpose of studying with Pearson, whose *Grammar of Science* had served as an inspiration since Neyman discovered it in 1916.

Since Neyman's 1923/1925 paper did not represent a conceptual breakthrough in the theory of sampling from finite populations, we are left with two questions of historical interest. The first is to ask how the isolated scholar arrived at what may well be an independent derivation of some key results in finite sampling and the second is to ask why Neyman's results, in particular, seem to have had a more profound influence than did similar ones obtained by others.

Perhaps the answer to the first question is that, while surely out of the mainstream of early twentieth century statistics, neither in Russia nor in Poland was Neyman completely isolated. Neyman studied probability theory in Kharkov under the direction of the Russian mathematician S.N. Bernstein as early as 1915 or 1916. Although these activities had been forgotten by Neyman by the time he was describing his early career to Constance Reid in 1978 (and perhaps this forgetfulness in his later life contributes to our image of Neyman's early years as totally isolated), in the mid 1920s Neyman described himself as having continued his studies under Bernstein through 1921. He also notes that he worked in 1920 under Bernstein to apply the theory of probability to experimentation in agriculture (Reid, 1982, p. 30). Through Bernstein, Neyman became familiar with Markov's work, which also influenced Tchouproff (see also Seneta, 1982, 1985). We know that, before he left Poland, Neyman worked with a Professor Isseroff on statistical analysis of agricultural experiments and even lectured to the Agricultural Department at the University of Kharkov on the application of probability theory to experimental problems in agriculture. These activities were clearly precursors to the papers dealing with experimental design in agriculture and his paper on finite sampling theory which appeared in *Biometrika* in 1925.

Just as Fisher was stimulated by the challenges of real experimentation at Rothamsted to gather together ideas that were "in the air" to make his masterly synthesis of the design of experiments and the analysis of variance, so Neyman could well have been stimulated by the challenges of experimentation at the Institute in Poland to synthesize ideas of sampling from finite populations that were "in the air". And that would lead rather naturally, given the variable citation practices at the time, to the kind of papers Neyman wrote in 1923, papers that had little reference to the work of others. The lack of references to work in English is particularly understandable, since as his advisor certified to Pearson, Neyman had no access to English journals while he was in Poland.

To answer the second question, that of Neyman's profound influence, we need to look not at the 1923 paper (nor at its 1925 republication), but at Neyman's watershed 1934 paper, in which he was able to capture the essential ingredients of the problem of sampling, synthesize his own contributions and those of others, and effectively demolish the idea of purposive sampling.

Neyman arrived in London in 1925, shortly after the publication of Fisher's *Statistical Methods for Research Workers*. But the paths of the two seem not to have crossed during the entire academic year. In March of 1926, Gosset passed through London and met with Neyman, who expressed an interest in visiting Fisher at Rothamsted. Gosset wrote to Fisher warning him that Neyman "holds a letter from me to you asking you to show him anything that you think might be useful to him. I daresay offprints would be useful to him if you could spare them as he finds it hard to trace your work. I gather that he will write to you in April". Fisher and Neyman actually met for the first time in July in Rothamsted, although there is no record of what they discussed. Neyman then spent a year in Paris, after which he returned to Poland. The next record of interaction between Neyman and Fisher is linked to Neyman's work in 1932 with Egon Pearson on the theory of statistical tests.

5 The 1925 and 1927 ISI Discussions on Sampling

Before Neyman returned to the problems of sampling from a finite population that he had addressed in 1923, major developments in sampling were presented at the 1925 and 1927 sessions of the International Statistical Institute (ISI), and published in 1926 and 1928, respectively. Kruskal & Mosteller (1980) give a related and somewhat more detailed discussion.

A substantial portion of the 1925 ISI Meeting was taken up with discussions of the "method of representative sampling". In 1924, the Bureau of the ISI appointed a Commission, consisting of Arthur Bowley, Corrado Gini, Adolph Jensen, Lucien March, Verijn Stuart, and Frantz Zizek, to study the representative method. Jensen served as rapporteur and leader of the discussion at the meeting. The Commission Report (Jensen, 1926a) contains a description of two methods: random sampling (with all elements of the population having the same probability of selection), and purposive selection of large groups of units (in modern terminology clusters) chosen to match the population on selected control variates. The report does not really attempt to choose between the methods. In one of several annexes to the report, Bowley (1926) provided a lengthy theoretical development, but failed to describe fully how the statistical theory of purposive sampling works. What is remarkable to us in this Report and its Annexes, especially in light of the controversy that Isserlis and Greenwood were to ignite only the next year, is the singular lack of references to the theoretical work on sampling from finite populations by Isserlis, Neyman, and Tchouproff, although the reference list in Jensen's (1926b) Annex did include one reference to Tchouproff, his 1910 thesis! Tchouproff was present at the meeting but because he was rather ill he appeared to contribute little to the discussion. He died soon after the meeting.

Yates (1946), in reviewing these proceedings, notes the "lack of any clear conception of the possibility, except by the selection of units wholly at random, or by the inadequate procedure of sub-dividing the sample into two or more parts, of so designing sampling inquiries that the sampling errors should be capable of exact estimation from the results of the inquiry itself". (p. 13). He goes on to describe the developments in random sampling that took place in England linked to agricultural experimentation between 1925 and 1935, stimulated largely by the elements of randomization in experimentation and the analysis of variance, both of which he attributes to Fisher. We see this as another indication of the lack of impact of Neyman's early contributions in these two interrelated fields.

At the 1927 ISI meeting, Corrado Gini presented a paper on the application of the purposive method to the sampling of records from the 1921 Italian census (see Gini, 1928, and Gini & Galvani, 1929), which was to play a pivotal role in the later work by Neyman. They needed to discard most of the records of the 1921 census before taking the next one, and they proposed to retain a sample for future analyses and reference. To make their sample representative, they chose to retain all of the data from 29 out of 214 large administrative districts into which Italy was divided. They applied the purposive method in order to select the 29 through a process which attempted to match the averages on seven important characteristics with those for the country as a whole. When they did this, they discovered that there were substantial deviations between the sample and the entire country for other characteristics. This, they claimed, called into question the accuracy of sampling. It remained for Neyman to take up the challenge embodied in this claim.

6 Neyman's 1934 Paper on Sampling

Neyman took up the challenge in his classic 1934 paper presented before the Royal Statistical Society, "On the two different aspects of the representative method". We review the paper's elements and emphasize how Neyman rescued clustering from the clutches of purposive sampling and gave it a rightful place in the foundations of random sampling methodology.

Neyman originally prepared the paper in 1932 in Polish (with an English summary) as a booklet growing out of his practical experience. He had been working for the Institute for Social Problems on a project involving sampling from the Polish census to obtain data to describe the structure of the working class in Poland. As he wrote to Egon Pearson, he used the opportunity and "pushed a little the theory" (Reid, 1982, p. 105). Published in 1933, the original Polish version of the paper traces a good deal of history (Neyman had clearly learned his lesson about citation practices and was now unquestionably familiar with the literature); the 1934 version published in the *Journal of the Royal Statistical Society* includes even more history and carefully cites many of the people who were to be in the room for the presentation before the Society. Neyman gives enormous credit to Bowley for his 1926 synthesis of the theory underpinning sampling methods, but also notes the hundreds of contributions in the area on which his comparison between purposive and random sampling builds. There is, however, the curious presentation of optimal allocation with no reference back to Tchouproff (1923a, 1923b).

The goal of the paper is to compare purposive and random sampling. But elements of synthesis are prominent as well. Neyman describes stratified sampling, noting that Bowley considers only the proportionate case but stating that such restriction is not necessary. He gives a crisp description of cluster sampling: "Suppose that the population $\prod$ of M' individuals is grouped into M_0 groups. Instead of considering the population $\prod$ we may now consider another population, say π, having for its elements the M_0 groups of individuals, into which the population $\prod$ is divided .... If there are enormous difficulties in sampling individuals at random, these difficulties may be greatly diminished when we adopt groups as the elements of sampling" (pp. 568–569). This is a new synthesis—earlier conceptualizations of clustering had coupled it with purposive sampling. Indeed, Neyman quotes Bowley as maintaining that "in purposive selection the unit is an aggregate, such as a whole district, and the sample is an aggregate of these aggregates, while in random selection the unit is a person or thing, which may or may not possess an attribute, or with which some measurable quantity is associated" (p. 570). Neyman goes on to explicitly uncouple clustering and purposive sampling, saying, "In fact the circumstance that the elements of sampling are not human individuals, but groups of these individuals, does not necessarily involve a negation of the randomness of the sampling" (p. 571). He calls this procedure "random sampling by groups" and points out that, although Bowley did not consider it theoretically, he used it in practice in London, as did O. Anderson in Bulgaria.

Neyman also speaks of combining stratification with clustering to form "random stratified sampling by groups". [Bowley, in his role as the lead discussant of Neyman's paper, notes the innovativeness of Neyman's suggestion of random stratified sampling of groups and acknowledges that in fact it was what he had been driven to use in his work even though he has not fully recognized the implications of what he had done]. Then Neyman refers to the full theory for best linear unbiased estimation[3] of a population average developed in his 1933 Polish publication, and he gives the now familiar formula for the variance of the "natural" weighted average estimator in stratified cluster sampling.

In order to compare Gini and Galvani's method of purposive selection with his own method of random stratified sampling by groups, Neyman imbues the purposive method with a structure that allows one to treat it as if it were based on a special form of random selection, even though this was clearly not the way Gini and Galvani actually selected or conceived of their groups (districts)[4].

[3]There was an implicit assumption in Neyman's work regarding the uniqueness of the best linear unbiased estimate in sampling from finite populations. This assumption turns out to be false. Godambe & Thompson (1971, pp. 386–387) observe that "Neyman proposed the use of the Gauss-Markoff technique to prove the optimality (UMV-ness) of the sample mean and other similar estimators. Indeed, during the discussion, Fisher concurred in this method; and it appears that the Gauss–Markoff technique, now known to be of doubtful validity in sampling theory (Godambe, 1955), was one of the very few points on which Neyman and Fisher were in agreement".

[4]Neyman envisions a universe of districts divided up into strata according to the values of one or more control variates, and then each stratum is subdivided into substrata according to the number of units in the district. Then within each substratum he speaks of the random selection of a preassigned number of districts. Of course, because of the large size of the districts in Gini and Galvani's situation, most strata would contain zero or one district. This leads to the sampling bias which Neyman then illustrates.

In doing so, Neyman constructs a sturdy coffin in which to bury the method of purposive selection, and then he proceeds to slam the lid of the coffin tight and nail it shut, by presenting one statistical argument after another. First, he considers the conditions under which the purposive method, which utilizes the regression of group means on a control variate y, produces unbiased estimates. He then notes, on the basis of calculations from Gini and Galvani's own data, that the conditions appear not to be satisfied in practice. Neyman goes further, however, and points out that Gini and Galvani make the implicit assumption that the groups (clusters) are themselves random samples from the population, something that is decidedly false. Neyman's own method of random stratified sampling by groups does not have these deficiencies. Neyman carries the argument one final step by exploring, using his variance formula, whether it is preferable to sample a small number of large groups (in effect a variant on Gini and Galvani) or a large number of smaller units. The latter turns out to be the clear choice and this provides Neyman with the final nail to hammer shut the coffin of the purposive method. Shortly later, the coffin was buried, although the ghost of the purposive method continues to rise until this day in the form of quota sampling.

The immediate effect of Neyman's paper was to establish the primacy of the method of stratified random sampling over the method of purposive selection, something that was left in doubt by the 1925 ISI presentations by Jensen and Bowley. But the paper's longer-term importance for sampling was the consequence of Neyman's wisdom in rescuing clustering from those who were the advocates of purposive sampling and integrating it with stratification in a synthesis that laid the groundwork for modern-day multistage probability sampling.

The heart of the 1934 paper, however, in terms of the amount of space Neyman devoted to exposition and in terms of emphasis in the discussion by others at the presentation before the Royal Statistical Society, is the material on confidence intervals. There was much confusion as to whether this was just Fisher's fiducial method presented in a slightly different fashion or a new inferential approach. Neyman introduced the idea of coverage in repeated samples, but the discussion did not pick up on its originality. It is somewhat ironic that in his discussion of Neyman (1934), Leon Isserlis, who only seven years earlier had condemned Neyman for not giving enough credit to Tchouproff, completely overlooked the fact that Neyman failed to credit Tchouproff in this paper, this time for proposing the notion of optimal allocation in stratified sampling. We speculate that Isserlis' failure to question this point arose largely from the trouble he was having understanding Neyman's concept of confidence intervals, the latter being the sole topic of Isserlis' discussion.

Two decades later Neyman (1952a) acknowledged Tschouproff's priority in discovering these results:

> "I am obliged to Dr. Donovan J. Thompson of the Statistical Laboratory, Iowa State College, Ames, Iowa, for calling my attention to the article of A.A. Tschuprow, "On the mathematical expectation of the moments of frequency distributions in the case of correlated observations" published in *Metron*, Vol. 2, No. 4 (1923), pp. 646–680, which contains some results refound by me and published, without reference to Tschuprow, in 1933.
>
> The results in question are the general formula for the variance of the estimate of a mean in stratified sampling and the formula determining the optimum stratification of the sample. These formulae appeared first in a Polish booklet *An Outline of the Theory and Practice of Representative Method, Applied in Social Research* published in 1933 by the Warsaw Institute of Social Problems. Later on they were republished in English in the *Journal of the Royal Statistical Society*, Vol. 97 (1934), pp. 558–625. Finally, the same formulae, again without a reference to Professor Tschuprow, were given in the second edition of my book, *Lectures and Conferences on Mathematical Statistics and Probability*, Washington, D.C., 1952.
>
> The purpose of this note is, then, to recognize the priority of Professor Tschuprow,

to express my regret for overlooking his results and to thank Dr. Thompson for calling my attention to the oversight."

And in the 1970s, Neyman continued in this apologetic vein by reprinting the recognition of priority in an introduction to a book of correspondence between Markov and Tschouproff (Neyman, 1981). What seems clear today for those of us who go back for Tschouproff (1923a, 1923b) is that if Neyman had read these papers, it would have been difficult to have missed Tschouproff's treatment of stratification although we confess to having to search to identify the result on optimal allocation. Thus, the simplest interpretation for us is that when Neyman was developing the material for the 1934 paper in Poland he not only did not have easy access to Tschouproff's paper, but simply presumed that everyone would accept the notion of stratification, especially in light of the ISI discussion and publications on the topic.

Fisher's lenghty discussion of the Neyman paper is noteworthy in the present context, not so much for his challenge that Neyman's confidence intervals were in some senses fiducial intervals under a different name, but rather for his observations on the importance of the parallelism between sampling in economic research and the role of sampling of finite populations in agricultural experimentation. The major distinction, he was reported as saying, was that: "In a well-designed experiment, however, the mathematics were simplified, and all anxiety was avoided in respect to different systems of weighting" (p. 616).

Of course, Fisher did recognize that Neyman's goal in creating confidence intervals was quite different from his own goal of inductive inferences from the sample at hand. As Edwards (1995) reminds us, Fisher expessed his skepticism regarding the value of the coverage property of confidence intervals and he remarked in his discussion that Neyman's generalization "was wide and handsome, but it had been erected at considerable expense, and it was perhaps as well to count the cost" (p. 618).

Our question of why the Neyman paper had such a profound influence compared to the earlier work of Tchouproff and others, was also raised by Bellhouse (1988) and Kruskal & Mosteller (1980). We believe that while Tchouproff had clearly derived a number of the technical results a decade earlier, and had dropped the constraint of constant probabilities of selection, his papers were abstract and formal in nature, and his results were far removed from real-world application. Neyman more clearly laid the groundwork for statistical practice by his innovative integration of clustering and stratification, and his clear and convincing exposition of the inferiority of the purposive method. Neyman provided the recipe for others to follow and he continued to explain its use in convincing detail to those who were eager to make random sampling a standard diet for practical consumption (e.g., see Neyman, 1952b, for a description based on his 1937 lectures on the topic at the U.S. Department of Agriculture Graduate School).

7 Fisher, Neyman, and The Design of Experiments in 1935

The year after Neyman presented his sampling paper to the Royal Statistical Society, three publications resolved all doubt about who was to gain full credit for the design of experiments as an area of statistics. To understand these events surrounding these publications, however, we move back several years and trace the interlocking developments in this area involving Fisher and Neyman.

Following Fisher's series of papers on experimental design in the 1920s, those at Rothamsted such as Yates and Wishart helped to propagate his ideas and move them into actual agricultural practice, and Fisher temporarily turned his focus to the completion of *The Genetical Theory of Natural Selection*, published in 1930, as well as related issues of genetics and eugenics and his ideas on inverse probability. But Fisher and others recognized the need to pull together the new ideas on statistics and experimentation. In 1930 he lectured on the topic at Imperial College in London and at Chelsea Polytechnic and then he spent the summer of 1931 at Iowa State University giving some related lectures. This was to be the beginning of notes that ultimately led to *The Design of Experiments*, in 1935. Meanwhile, Neyman was working on a series of papers on hypothesis testing with Egon Pearson, and on some agricultural experimentation problems.

By the late 1920s, Fisher was becoming restless at Rothamsted and considered applying for positions at British universities (Box, 1978, pp. 202–203). Box (1978) and Reid (1982) describe some of the correspondence between Fisher and Neyman in 1932 and 1933, primarily about Neyman's paper with Egon Pearson on tests of hypotheses. When Karl Pearson retired from University College in 1933, both his position and the department were split in two, with Egon Pearson being appointed as Reader in Statistics and head of the Department of Applied Statistics and Fisher being given the position of Galton Professor of Eugenics and responsibility for the Galton Laboratory. A major issue between the two was responsibility for teaching of statistics. Pearson proposed that Fisher not teach statistical theory, and so Fisher responded by proposing to teach the logic and philosophy of experimentation (Box, 1978, p. 259 and Bennett, 1990, p. 192). Here we see another impetus for Fisher to complete the book on experimental design. In mid 1933, upon hearing of Fisher's appointment at University College, Neyman wrote to him asking about the possibility of a position in the Laboratory. But Fisher had nothing to offer specifically in statistics and, in 1934, Neyman left Poland and joined Pearson's department for a three month visiting position, which ultimately turned into a permanent one. It was also at about this time that Fisher nominated Neyman for membership in the International Statistical Institute. That first term back in London, Neyman lectured on interval estimation and Fisher on experimentation, with Neyman in attendance.

In March of 1935, Neyman presented a paper before the recently formed Industrial and Agricultural Research Section of the Royal Statistical Society (Neyman, 1935) entitled "Statistical problems in agricultural experimentation". The paper was based on work done in Poland, in co-operation with two junior colleagues, Iwaszkiewicz and Kolodziejczyk. In it, he presented formal statistical models for the analysis of randomized block and Latin square designs, and he claimed that "the application of the usual tests of significance to the results of experiments by the two methods is not yet rigorously justified, though the effect of the inaccuracy involved may be negligible". In other words (but not those explicitly used by Neyman), Fisher's claims about the validity of tests were wrong! Neyman gives the model explicitly; treatments affect individual plots in a row-by-column layout differently (i.e., unit-treatment non-additivity). The model and Neyman's discussion of it includes the notion of an infinite series of repetitions of the experiment, and it is with respect to this long-run distribution that he takes expectations. In an appendix, he shows that, if the average treatment effect over the entire experimental area is the same for all treatments, the usual estimate of the error variance is unbiased for randomized blocks whereas for his Latin square model there is a bias. He repeats the opening claim that this bias may not be substantial. This means that the F test (referred to by both Fisher and Neyman as the z test) for treatment effects in a Latin square "may ... cause a tendency to state significant differentiation when this, in fact, does not exist" (p. 154). Then Neyman goes on to examine the test of hypothesis that one treatment has a greater effect than another and he demonstrates, in a series of examples, that Latin squares are often less efficient than randomized blocks. Throughout the paper Neyman employed the notions of type I and type II errors he had developed with Pearson.

Fisher was the lead discussant of the paper and he responded with great sarcasm about Neyman's surprising results noting that those present "had to thank him, not only for bringing these discoveries to their notice, but also for concealing them from public knowledge until such time as the method should be widely adopted in practice!" (p.155). Fisher questioned Neyman's apparent inability to grasp the very simple argument by which the unbiased character of the test of significance might be demonstrated and then he presented a simple illustration in the context of a 5×5 Latin square. Here Fisher states rather obliquely one of his key differences with Neyman, the hypothesis to be tested, which in the Fisherian formulation is that treatments have no effect on yields whereas in the

Neyman formulation it relates to zero treatment effects averaged over all plots. Then Fisher invokes the randomization distribution, which in effect assumes unit-treatment additivity. Yates and Wishart also contributed to the discussion in a more polite, but just as critical fashion, and when Neyman attempted to respond briefly, Fisher interrupted with the final statement:

> "I think it is clear to everyone present that Dr. Neyman has misunderstood the intention—clearly and frequently stated—of the z test and of the Latin Square and other techniques designed to be used with that test. Dr. Neyman thinks that another test would be more important. I am not going to argue that point. It may be that the question that Dr. Neyman thinks should be answered is more important than the one I have proposed and attempted to answer. I suggest that before criticizing previous work it is always wise to give enough study to the subject to understand its purpose. Failing that it is surely quite unusual to claim to understand the purpose of previous work better than its author.
>
> From a purely statistical standpoint the position is that Dr. Neyman would like a different test to be made, and I hope he will invent a test of significance, and a method of experimentation, which will be as accurate for questions he considers to be important as the Latin square is for the purpose for which it was designed."

Neyman later responded at length in writing but his relation with Fisher had been shattered and most present at the presentation of the paper seemed to side with Fisher. The controversy that began with this exchange raged throughout Fisher's lifetime, and it included not only fundamental aspects of experimental design but also two very different approaches to testimony and statistical inference.

The dispute that erupted over the Neyman paper was bound up in issues of estimation and hypothesis testing and the difference between the two types of inference. Even with Neyman's model, the usual F-test in the Fisherian approach has the correct null distribution when there is "no treatment effect", because then there is no issue of whether treatment-unit additivity holds. Fisher correctly pointed this out in the discussion. Neyman was insisting on a different null hypothesis, that the treatment effect, through its interaction with plots, averages out over the complete set of plots used in the experiment. This is essentially estimating the treatment effect in the non-null case, however, and here it matters whether one allows for treatment-unit additivity or not. Fisher did allude to this difference, but in language that remains difficult to decipher. (Rhetoric was at a premium rather than statistical clarity.) Nelder (1994) essentially argues that Neyman's error was in formulating an uninteresting null hypothesis, largely as a result of the non-hierarchical nature of its construction. In such circumstances one also needs to look at the alternatives. Thus Nelder argues that Fisher's attack on Neyman was irrelevant because he dealt only with the null case. Further, the Neyman/Pearson approach to confidence intervals and tests of hypotheses, which Fisher was never to accept, was essential to the Neyman argument in this paper; the ideas were still too new for most present to understand. Finally, we note that randomization seemed to play little role in Neyman's arguments in the 1935 paper whereas it was essential to Fisher's ideas, clearly structuring all of his thinking on models and the validity of related statistical methods for the analysis of experiments.

Just two months after Neyman's presentation, Fisher's successor at Rothamsted, Frank Yates (1935), presented his own paper on agricultural experimentation before the same section of the Royal Statistical Society and the debate between Fisher and Neyman continued, with Yates becoming the target for Neyman's criticism. In a lengthy critique of Yates' exposition of the theory of factorial experimentation, submitted in writing after the meeting, Neyman again used his work with Pearson on tests of hypotheses to suggest that interactions had low power of detection and that Yates' (and thus Fisher's) definition of interactions were fatally flawed. This discussion seems to have confused the nature of the contrasts used to estimate effects in the Fisher–Yates framework and Yates avoided the possibility of others being similarly confused by altering slightly the definitions in the printed version of his paper (which followed Neyman's paper in the *Journal*). This minor alteration vitiated all of Neyman's analysis. Yates also responded rather caustically in writing to Neyman's critique and Neyman then submitted an addendum pointing out why he still thought the Fisher–Yates approach was wrong. The controversy between Neyman and Fisher (with Yates as an occasional surrogate) was now full-blown.

The final of the trio of 1935 publications on experimentation, Fisher's *The Design of Experiments*, was the most significant. Fisher completed the preface the month after Yates presented his paper. In this book, Fisher presented an integrated perspective on the logic and principles of experimentation. After a brief introduction in which he explains why for inferences he does not use Bayes' theorem in the book, Fisher turns to his famous example of "The Lady Tasting Tea", based on an actual experiment he proposed more than 12 years earlier at Rothamsted, in response to the claim by Muriel Bristol, an algologist, who claimed that she could tell the difference between a cup of tea into which the milk had been poured first (her preference) and one into which the tea had been poured first (which Fisher had offered her). Here Fisher made his strongest arguments for the need to randomize. In the next chapter he dealt with Darwin's data on cross- and self-fertilization (in which there was no randomization), introducing the notion of matching, and this time he used the argument of randomization to "justify" the use of the t-distribution for the key comparison. Subsequent chapters dealt with randomized blocks, Latin squares, factorial designs, confounding, partial confounding, and the analysis of covariance. The final pair of chapters dealt with more theoretical material on estimation: fiducial inference and the measurement of information.

Few of the ideas in the book were totally new, but their integration and the power of Fisher's exposition of them helped to establish randomized experimentation as basic to the scientist's toolkit. The battle over randomization continued to be waged in British journals, with Gosset, who still advocated the use of systematic designs such as Beavan's drill strip or sandwich design, being Fisher's main antagonist. The debate essentially ended with Gosset's death in 1937, although a final critique by Gosset appeared posthumously in *Biometrika* in 1938. On the other side of the Atlantic, there was no such battle, and in 1936, on the occasion of its 300th anniversary, Harvard University awarded Fisher an honorary degree at least in part for this achievement. Just as with Neyman and his 1934 paper on sampling, Fisher's *The Design of Experiments* provided a recipe for others to follow.

8 Conclusions

Both Fisher and Neyman made fundamental contributions to statistical methodology that underpins the modern approaches to the design of experiments and the design of sample surveys. In this paper we have retraced their work on these topics, especially during the critical years from 1921 to 1935, and we attempt to explain the pivotal roles of Neyman's 1934 paper on agricultural experimentation which ultimately led to a rift between the two men that was never to be repaired. In the aftermath, Neyman staked out intellectual responsibility for sampling and Fisher did the same for experimentation. The two fields subsequently drifted apart.

Two issues were raised in the discussion of an earlier version of this paper presented at the 50th ISI Session in Beijing, China. In this discussion, others challenged us regarding the real reason for the separation of the design of experiments and sample surveys after 1935, and regarding the original of Neyman's contributions to sampling. We comment on each briefly.

Some have argued that it was the fundamental differences between experimentation and survey work that were truly responsible for the separation of the two fields within statistics and not the fued between Fisher and Neyman. For example, T.F.M. Smith (personal communication) argues that randomization in an experiment supports internal validity, to the collection of experimental units, whereas random selection in a survey setting supports external validity, allowing generalization to those not in the sample. This is technically correct, but only in a narrow sense. The statistical tools used in the two fields are essentially the same and the real scientific goals of an experimenter involve generalization beyond the experimental units. This is exactly why we have emphasized the importance of experiments embedded in surveys and surveys embedded in experiments (Fienberg & Tanur, 1988, 1989). The recent interest in the common features of experimentation and sampling suggests that the technical links are important, as many of the early contributors recognized in the 1930s. Thus we do not believe that the differences in perspective between experimentation as surveys *alone* can account for the separation. The clash of two strong personalities who were determined to gain recognition for their inferential approaches, in our view, was a major contribution to the split.

Others continue to raise the issue of the extent to which Neyman "stood on the shoulders of giants". In particular, Chris Heyde (personal communication) raises the question of Neyman's debt to the Russian school of sampling led by Tschouproff. It is true that Neyman was educated in Karkov and studied probability with Bernstein, as we note above, but he appears to have had little or no contact with statistics and the sampling of finite populations. Thus, as we argued above, we believe that the contributions in Neyman's 1923 paper represent independent discovery. In the 1925 Biometrika version there is a reference to an earlier Biometrika paper by Tschouproff, but no recognition of Tschouproff's 1923 Metron paper. This appears somewhat sloppy and even in 1934, Neyman did not single out Tschouproff for recognition, either in general or in connection with optimal allocation. But Neyman's purpose in 1934 was different and he brought a new integrative perspective to sampling in the paper that differed from those who wrote on the topic previously.

Thus while we believe that Neyman was influenced by the work of many others—few of whom he referenced—he brought originality to his papers on sampling and the clarity of his exposition was crucial to the subsequent development of sampling. As a result, he rightly became one of the giants on whose shoulders others have stood.

We continue to see important links between the fields of experimentation and sample surveys and we find their roots in the pioneering work of both Fisher and Neyman, during a time of great intellectual ferment in the 1920s and 1930s. We believe that the statistical profession has much to learn from their creative ideas.

Acknowledgments

A preliminary version of parts of this paper dealing with Neyman was presented at the "International Conference on Statistics to Commemorate the 100th Anniversary of Jerzy Neyman's Birth", Jarchranka, Poland, November 25–26, 1994, and appeared as Fienberg & Tanur (1995a). The present paper has added considerably to the material drawn from that earlier one. We thank A.W.F. Edwards, Chris Heyde, John Nelder, Eugene Seneta and Fred Smith for comments on an earlier version of the present paper, which was presented at the 50th ISI Session in Beijing, China in August 1995 (Fienberg & Tanur 1995b).

References

Bellhouse, D.R. (1988). A brief history of random sampling methods. In *Handbook of Statistics*. Eds. P.R. Krishnaiah & C.R. Rao, Vol. 6. North Holland, Amsterdam, pp. 1–14.

Bowley, A.L. (1926). Measurement of the precision attained in sampling. (Annex A to the Report by Jensen). *Bulletin of the International Statistical Institute*, **22**, Supplement to Liv. 1: 1–62.

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

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

Box, J.F. (1980). R.A. Fisher and the Design of Experiments, 1922–1926. *American Statistician*, **34**, 1–7.

Church, A.E.R. (1926). On the moments of the distribution of squared standard deviations of small samples from any population. *Biometrika*, **18**, 321–394.

Cochran, W.G. (1980). Fisher and the analysis of variance. In *R.A. Fisher: An Appreciation*. Eds. S.E. Fienberg and D.V. Hinkley. Springer-Verlag: New York, pp. 17–34.

Eden, T. & Fisher, R.A. (1927). Studies in crop variation. IV. The experimental determination of the value of top dressings with cereals. *Journal of Agricultural Science*, **17**, 548–562.

Edwards, A.W.F. (1995). XVIII-th Fisher Memorial Lecture delivered at the National History Museum, London on Thursday, 20th October, 1994: Fiducial Inference and the Fundamental Theorem of Natural Selection. *Biometrics*, **51**, 799–809.

Fienberg, S.E. & Tanur, J.M. (1987). Experimental and sampling structures: Parallels diverging and meeting. *International Statistical Review*, **55**, 75–96.

Fienberg, S.E. & Tanur, J.M. (1988). From the inside out and the outside in: Combining experimental and sampling structures. *Canadian Journal of Statistics*, **16**, 135–151.

Fienberg, S.E. & Tanur, J.M. (1989). Combining cognitive and statistical approaches to survey design. *Science*, **243**, 1017–1022.

Fienberg, S.E. & Tanur, J.M. (1995a). Reconsidering Neyman on experimentation and sampling: Controversies and fundamental contributions. *Probability and Mathematical Statistics*, **15**, 47–60.

Fienberg, S.E. & Tanur, J.M. (1995b). Reconsidering the fundamental contributions of Fisher and Neyman on experimentation and sampling. *Bulletin of the International Statistical Institute, 50th Session*, Invited Papers Book 3, 1243–1260.

Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, **52**, 399–433.

Fisher, R.A. (1921). Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *Journal of Agricultural Science*, **11**, 107–135.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fisher, R.A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture*, **33**, 503–513.

Fisher, R.A. (1935). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Fisher, R.A. (1952). Statistical methods in genetics. *Heredity*, **6**, 1–12.

Fisher, R.A. & Eden, T. (1927). Studies in crop variation. II. The experimental determination of the value of top dressings with cereals. *Journal of Agricultural Science*, **17**, 548–562.

Fisher, R.A. & Mackenzie, W.A. (1923). Studies in crop variation. II. The manurial response of different potato varieties. *Journal of Agricultural Science*, **13**, 311–320.

Gini, C. (1928). Une application de la méthode représentative aux matériaux du dernier recensement de la population italienne (1er décembre 1921). *Bulletin of the International Statistical Institute*, **23, Liv. 2**, 198–215.

Gini, C. & Galvani, L. (1929). Di una applicazione del metodo rappresentative all'ultimo censimento Italiano della popolazione (1° decembri, 1921). *Annali di Statistica*, **Series 6**, 4, 1–107.

Godambe, V.P. (1955). A unified theory of sampling from finite population. *Journal of the Royal Statistical Society, Series B*, **17**, 269–278.

Godambe, V.P. & Thompson, M.E. (1971). Bayes, fiducial and frequency aspects of statistical inference in regression analysis in survey sampling (with discussion). *Journal of the Royal Statistical Society, Series B*, **17**, 361–390.

Greenwood, M. & Isserlis, L. (1927). A historical note on the problem of small samples. *Journal of the Royal Statistical Society*, **90**, 347–352.

Isserlis, L. (1918). Formulae for determining the mean values of products of deviations of mixed moment coefficients in two to eight variables in samples taken from a limited population. *Biometrika*, **12**, 183–184.

Jensen, A. (1926a). Report on the representative method in statistics. *Bulletin of the International Statistical Institute*, **22, Liv. 1**, 359–380.

Jensen, A. (1926b). The representative method in practice. (Annex B to the Report by Jensen). *Bulletin of the International Statistical Institute*, **22, Liv. 1**, 381–439.

Kruskal, W.H. & Mosteller, F. (1980). Representative sampling, IV: The history of the concept in statistics, 1895–1939. *International Statistical Review*, **48**, 169–195.

Lehmann, E.L. & Reid, C. (1982). In Memoriam, Jerzy Neyman 1894–1981. *American Statistician*, **36**, 161–162.

Nelder, J.A. (1994). The statistics of linear models: back to basics. *Statistics and Computing*, **4**, 221–234.

Neyman, J. (1933). *Zarys teorii i praktyki bandania struktury ludnosci metoda reprezentacyjna (An Outline of the Theory and Practice of Representative Method, Applied in Social Research)*. Institute for Social Problems, Warsaw. (In Polish with an English summary).

Neyman, J. (1934). On 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**, 558–625.

Neyman, J. (1952a). Recognition of priority. *Journal of the Royal Statistical Society*, **115**, 602.

Neyman, J. (1952b [1938]). *Lectures and Conferences on Mathematical Statistics and Probability*. U.S. Department of Agriculture, Washington, DC. (The 1952 edition is an expanded and revised version of the original 1938 mimeographed edition).

Neyman, J. (1981). Introduction to *The Correspondence Between A.A. Markov and A.A. Chuprov on the Theory of Probability and Mathematical Statistics*, (Kh. O. Ondar, ed.), Springer-Verlag, New York.

Neyman, J., with the cooperation of K. Iwaszkiewicz and St. Kolodziejczyk. (1935). Statistical problems in agricultural experimentation (with discussion). *Supplement to the Journal of the Royal Statistical Society*, **2**, 107–180.

Pearson, K. (1927). Another "historical note on the problem of small samples". (Editorial) *Biometrika*, **19**, 207–210.

Reid, C. (1982). *Neyman from Life*. Springer-Verlag, New York.

Rubin, D.B. (1990). Comment: Neyman (1923) and causal inference in experiments and observational studies. *Statistical Science*, **5**, 472–480.

Russell, E.J. (1926). Field experiments: How they are made and what they are. *Journal of the Ministry of Agriculture*, **32**, 989–1001.

Seneta, E. (1982). Chuprov (or Tschuprow), Alexander Alexandrovich. In *Encyclopedia of Statistical Sciences*, Vol. 1 (S. Kotz and N. Johnson, eds.). Wiley, New York, 477–479.

Seneta, E. (1985). A sketch of the history of survey sampling in Russia. *Journal of the Royal Statistical Society, Series A*,

148, 118–125.

Splawa–Neyman, J. (1925[1923b]) Contributions of the theory of small samples drawn from a finite population. *Biometrika*, **17**, 472–479. (The note on this republication reads "These results with others were originally published in *La Revue Mensuelle de Statistique*, publ. parl'Office Central de Statistique de la République Polonaise, tom. vi. pp. 1–29, 1923").

Splawa–Neyman, J. (1990 [1923a]). On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, **5**, 465–472. Translated and edited by D.M. Dabrowska & T.P. Speed from the Polish original, which appeared in *Roczniki Nauk Rolniczyc*, **Tom X** (1923): 1–51 (*Annals of Agricultural Sciences*).

'Student' (1923). On testing varieties of cereals. *Biometrika*, **15**, 271–293.

'Student' (1938). Comparison between balanced and rand arrangements of field plots. *Biometrika*, **29**, 363–369.

Tschuprow, Al.A. (1923a). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations (Chapters I-III). *Metron*, **2**, 461–493.

Tschuprow, Al.A. (1923b). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations (Chapters IV-VI). *Metron*, **2**, 646–680.

Yates, F. (1935). Complex experimentation (with discussion). *Supplement to the Journal of the Royal Statistical Society*, **2**, 181–247.

Yates, F. (1946). A review of recent developments in sampling and sample surveys (with discussion). *Journal of the Royal Statistical Society*, **109**, 12–42.

Résumé

R.A. Fisher and Jerzy Neyman sont bien reconnus comme les statisticiens qui ont établi les idées fondamentales qui soutiennent le plan des expériences et le plan des enguêtes par sondage, respectivement. Dans cet article nous revoyons les contributions centrales de ces hommes fameux dans les deux domaines de recherche. Nous adressons aussi l'effet d'une controverse qui a rapport á l'article de Neyman (1935) au sujet de l'expérimentation agronome qui a abouti á une separation dans le champs des recherches des expériences et des échantillonnages.

[*Received March 1996, accepted July 1996*]