

## ON REREADING R. A. FISHER

BY LEONARD J. SAVAGE<sup>1,2</sup>

*Yale University*

[Fisher's contributions to statistics are surveyed. His background, skills, temperament, and style of thought and writing are sketched. His mathematical and methodological contributions are outlined. More attention is given to the technical concepts he introduced or emphasized, such as consistency, sufficiency, efficiency, information, and maximum likelihood. Still more attention is given to his conception and concepts of probability and inference, including likelihood, the fiducial argument, and hypothesis testing. Fisher is at once very near to and very far from modern statistical thought generally.]

### 1. Introduction.

1.1. *Why this essay?* Of course an R. A. Fisher Memorial Lecture need not be about R. A. Fisher himself, but the invitation to lecture in his honor set me so to thinking of Fisher's influence on my own statistical education that I could not tear myself away from the project of a somewhat personal review of his work.

My statistical mentors, Milton Friedman and W. Allen Wallis, held that Fisher's *Statistical Methods for Research Workers* (RW, 1925) was the serious

---

<sup>1</sup> This is a written version of an R. A. Fisher Memorial Lecture given in Detroit on 29 December 1970 under the auspices of the American Statistical Association, the Institute of Mathematical Statistics, and the Biometric Society. The author's untimely death on 1 November 1971 occurred before he had finished work on the paper. The editor for publication was John W. Pratt. The author's work on the lecture and his subsequent paper and the editor's work were supported in part by the Army, Navy, Air Force, and NASA under a contract administered by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

<sup>2</sup> Savage had drafted all the text given here, except the abstract. In the manuscript, the text through § 2.5 is thoroughly revised in his handwriting and the remainder not at all, though it may be a second draft through Section 4.3, where the typing changes, and a first draft thereafter. He left the references incomplete through § 2.5 and almost completely absent thereafter. He indicated where he intended to quote Fisher, but specified what he intended to quote only in § 1.1.

The editor therefore had to provide a majority of the references and all but the first quotation. All comments in parentheses with references and all footnotes were also added by the editor. These additions are not flagged in the printed text, nor are minor changes of no substance, including a few relocations of material. All other alterations of even minor substance are identified as 'insertions' or 'changes'. The changes usually replace material of similar length and  
(continued on next page)

*AMS 1970 subject classifications.* Primary 62A99; Secondary 62-03.

*Key words and phrases.* Statistical inference, R. A. Fisher, R. A. Fisher's contributions to statistics, estimation, consistency, sufficiency, efficiency, information, maximum likelihood, likelihood function, likelihood principle, hypothesis testing, fiducial probability, ancillary statistics, design of experiments, randomization, statistic, induction, inductive behavior, inverse probability, probability definition, reference set, Bayes' rule, sampling distributions, Fisher consistency, Behrens-Fisher problem,  $k$ -statistics, probit, randomization tests, Karl Pearson, Edgeworth.

man's introduction to statistics. They shared that idea with their own admired teacher, Harold Hotelling. They and some others, though of course not all, gave the same advice: "To become a statistician, practice statistics and mull Fisher over with patience, respect, and skepticism."

topic. The only deletion of any length or substance (in § 4.2) is footnoted. Most alterations of the text are for clarity or to bring in relevant ideas expressed elsewhere by Savage. A few amend or go beyond his intentions, but only on relatively objective matters where the evidence seems clear. The editor has tried to keep his personal views on controversial issues to the footnotes. Complete, unambiguous documentation would, however, have required excessive fussiness and footnoting. Those concerned with the nature and purpose of the editorial alterations will usually be able to deduce them, more easily than this description may suggest. Those who want all evidence on Savage's thought would need to consult the original materials in any case.

Material prepared by Savage in addition to the manuscript include about 200 index cards, some with more than one entry; about 50 handwritten pages, which the editor has had typed, of sometimes fascinating "random notes" on many works by Fisher and a few by others; Savage's personal copies, which he sometimes marked, of most of these works and quite a few more; and about 25 other pages of notes, mostly references and lists of topics. A tape of the original lecture was available and has been transcribed. All these materials were useful in the editing, especially for filling in references, but they by no means resolved all problems. They and Savage's other scientific papers, including correspondence, are available, excellently indexed, in archives at Sterling Memorial Library, Yale University.

The editor is grateful for help, especially with references not by Fisher, to the discussants, most of whom sent separate comments for editorial use; and in addition to the following, with apologies to anyone inadvertently omitted: F. J. Anscombe, G. A. Barnard, M. S. Bartlett, J. H. Bennett, R. J. Buehler, H. Chernoff, W. G. Cochran, A. P. Dempster, A. W. F. Edwards, D. J. Finney, J. Gurland, J. A. Hartigan, B. M. Hill, D. G. Kendall, M. G. Kendall, W. H. Kruskal, E. E. Leamer, L. Le Cam, E. L. Lehmann, F. Mosteller, E. S. Pearson, G.-C. Rota, I. R. Savage, H. Scheffé, E. L. Scott, and D. L. Wallace. That so many should have responded so generously is a tribute to Savage in itself. Of course this implicates them in no way.

Citations without dates, in the footnotes, are to these responses. Many interesting reactions could not be used, however. Eisenhart and Pearson, in particular, each wrote many pages of commentary of great interest, historical as well as substantive. All the responses are available in the archives.

Savage would obviously have revised his paper, especially the latter portions, considerably for style and some for substance. After circulating it for reaction, he would presumably have revised it further. In particular, offense would have been eliminated where he did not intend it but some now find it. An editor cannot know how he would have made any of these kinds of revisions, however, and any attempt risks distorting his meaning. Richard Savage and I therefore decided in the end to let the text stand even where it clearly presents a problem unless a resolution was also clear. We trust readers will make appropriate allowance for the unfinished state of the manuscript. Jimmie Savage once wrote (1954 vii):

One who so airs his opinions has serious misgivings that (as may be judged from other prefaces) he often tries to communicate along with his book. First, he longs to know, for reasons that are not altogether noble, whether he is really making a valuable contribution...

Again, what he has written is far from perfect, even to his biased eye. He has stopped revising and called the book finished, because one must sooner or later.

Finally he fears that he himself, and still more such public as he has, will forget that the book is tentative, that an author's most recent word need not be his last word.

In my first years as a teacher of statistics, I used Fisher's *Statistical Methods for Research Workers* several times as a textbook for an introductory graduate level course and also taught the theory of design from his *Design of Experiments* (DOE, 1935). It seems unlikely that Fisher's books will ever again be used as introductory textbooks, and even 20 years ago there was much to be said against it, but the experience was by no means all bad either for me or for the students.

The volume of Fisher's papers *Contributions to Mathematical Statistics* (CMS, 1950) appeared during the period of my most active work on *The Foundations of Statistics* (1954), upon which it had a great influence. Beginning about that time, I occasionally had the privilege of exchanging letters with Fisher, and much more rarely of talking with him.

Only figuratively can my preparation for this essay be referred to as rereading R. A. Fisher. I have long ago read considerably in the three books already mentioned, in Fisher's much more recent *Statistical Methods and Scientific Inference* (SI, 1956), and in a few other papers by him as well as in the extremely educational introduction to *Statistical Tables for Biological, Agricultural and Medical Research* (ST, Fisher and Yates, 1938). But I cannot pretend even now to have read all of his work or even to have read all that I explicitly refer to. His statistical corpus is so large and diverse that scarcely anyone but Fisher himself would be in a position to read even the papers in the 1950 collection (CMS) with comprehension, 'let alone all his *Collected Papers* (CP) and books.'

There are many ways in which such a rich body of writing might be reviewed. My aim here is to convey to you Fisher's main beliefs and attitudes—his viewpoint about statistics. There is a world of R. A. Fisher at once very near to and very far from the world of modern statisticians generally, and I hope to provide you with a rough map of it.

But what can be done for you in this that you cannot do for yourselves? Those who have already read in Fisher will agree that understanding him deeply is not easy, and they may be glad to hear the views of another. Some who have too long deferred plunging into Fisher will, I hope, find this essay a stimulating invitation.

As background, something will first be said about Fisher's interests and technical achievements, his manner and his relations with other statisticians. This will make more meaningful the review of his major ideas about statistical inference with special reference to the unusual, the unclear, and the controversial.

So many slippery topics must be touched on that no one is likely to agree with all my judgements, and at least a few objective errors are sure to resist eradication. If you find many doubts and few unequivocal answers do not be unduly disappointed. Fisher said 'once, in a sentence not characteristic of his writing':

I am still too often confronted by problems, even in my own research, to which I cannot confidently offer a solution, ever to be tempted to imply that finality has been

reached (or to take very seriously this claim when made by others!). (CMS, Preface)

1.2. *Sources of information.* The following pages consist largely of judgments and impressions. Nothing can convert these subjective reactions completely into objective facts, but it has been an invaluable discipline for me to support each of them by specific citations with reasonable thoroughness. These citations will also be useful to anyone who wants to reach his own conclusions about particular points, and I hope that they will not interfere with smooth reading of the text.<sup>3</sup>

To M. G. Kendall (1963) [and Bennett and Cornish (CP)] we owe a very complete bibliography of Fisher. That it is very long is no surprise, but its diversity may have surprised even Fisher's closest friends. [ *Statistical Methods for Research Workers* also contains a list of his statistical publications and some others.]

Among valuable works about Fisher are: Barnard 1963; Bartlett 1965, 1968; Hotelling 1951; M. G. Kendall 1963; Mahalanobis 1938; Neyman 1951, 1961, 1967; Pearson 1968, 1974; Yates and Mather 1963; and a forthcoming biography by his daughter, Joan Box. There is material of great value for biographical study of both Fisher and "Student" in a circulated, but not commercially published, collection of letters mainly from "Student" to Fisher (Gosset 1962).

## 2. Silhouette.

2.1. *Beginning before the beginning.* My central object is to delineate Fisher's outlook on statistics, but much of its importance and meaning would be lost were it not preceded by account of the man, his work, and style. Looking behind the scenes and reading between the lines are only human. These activities are both valuable and dangerous. Fisher himself illustrates excess in this sort of thing when he denigrates "mathematicians" as unfit because of their mathematical minds and training to comprehend the existence of nondeductive reasoning or the role of genetics in evolution (1932 257; 1935 b 39; 1935 f 155; 1936 a 248–9; 1958 a 261; he often stresses need for or lack of contact with Natural Sciences: RW ix; DOE 44; 1938; 1939 d 5–6; 1941 c 141; 1948 219; 1955 69; 1958 a 274; 1960 9; SI 102). Ironically, Karl Pearson, who was the most sinister of Fisher's "mathematicians" (1936 a 248–9; 1937 a esp. 302 a) was actually far less a mathematician than Fisher himself (1922 a 311; 1937 b esp. 311; SI 3; see also §§ 2.4 and 2.6). Again, George Pólya—a mathematician whose great work on combinatorics (1937) might greatly have enriched Fisher's own combinatoric work (1942 c, 1950 a) had Fisher known of it—has been actively interested for

<sup>3</sup> Where many citations are possible, Savage may have intended to be selective. In citing Fisher, rather than impose my own selection, I have tended to be overinclusive. In citing others than Fisher, on statistical topics generally, I have aimed to give a few helpful references, but not to be definitive. To minimize interference with smooth reading, I have used the most compact feasible style of citation. In particular, the same name applies to a string of dates until another name appears.

decades (1941, 1962–5) in the vital role of induction in mathematical activity. No less flagrant examples of misspent intellectualism [than this of Fisher's] can be found in many authors, including Fisher's critics.

*2.2. Background and skills.* Of course Fisher was not specifically trained to be a statistician. Only after Fisher was a great statistician, and largely because of the vision of statistics to which his activities gave rise, was statistical training inaugurated in a few universities.<sup>4</sup>

Fisher was a Cambridge-trained mathematician (see references in § 1.2), and despite what sometimes seems scorn for mathematicians, he was a very good one (Neyman 1951; 1961 147; 1967) with an extraordinary command of special functions (1915, 1921 a, 1922 c, 1925 b, 1925 c, 1928 a, 1931 b), combinatorics (1942 b, 1942 c, 1945 a, 1950 a, DOE, ST), and truly geometric  $n$ -dimensional geometry (1913, 1915, 1929 b, 1940 a; see also 1922 a, 1922 b, 1924 a, 1928 b, 1929 a, 1930 a, 1939 b). Indeed, my recent reading reveals Fisher as much more of a mathematician than I had previously recognized. I had been misled by his own attitude toward mathematicians, especially by his lack of comprehension of, and contempt for, modern abstract tendencies in mathematics (1958 a; see also § 2.1; yet see 1942 b esp. 340 a; 1945 a; DOE § 45.1). Seeing Fisher ignorant of those parts of mathematics in which I was best trained, I long suspected that his mastery of other parts had been exaggerated, but it now seems to me that statistics has never been served by a mathematician stronger in certain directions than Fisher was. No complete statistician is merely a mathematician, and Fisher—like other statisticians of his time—was a skilled and energetic desk calculator (RW examples; ST), tabulator (ST; see Index at “Tables” in RW and CMS or CP), and grapher (RW Ch. 2; 1922 a § 10; 1924 c; 1928 d). He early became a widely experienced and resourceful applied statistician, mainly in the fields of agronomy and laboratory biology (see his bibliography; the examples in RW and DOE; practical suggestions in 1926 a; in RW Ch. 2; and in DOE § 10 par. 2, § 12, § 25, § 29, end of § 60).

In addition to Fisher's illustrious career as a statistician he had one almost as illustrious as a population geneticist, so that quite apart from his work in statistics he was a famous, creative, and controversial geneticist (see references in § 1.2). Even today, I occasionally meet geneticists who ask me whether it is true that the great geneticist R. A. Fisher was also an important statistician. Fisher held two chairs in genetics, first at University College, London, and then at Cambridge, but was never a professor of statistics.

*2.3. Temperament.* Fisher burned even more than the rest of us, it seems to me, to be original, right, important, famous, and respected. And in enormous

<sup>4</sup> Advanced training in theoretical statistics and its application has been available at University College, London since the 1890's (Pearson 1974), but Savage's statement is surely correct in spirit, and technically as well if “training” means at the doctoral level and “a few” means more than one or two.

measure, he achieved all of that, though never enough to bring him peace. Not unnaturally, he was often involved in quarrels, and though he sometimes disagreed politely (1929f; 1929g; 1930c 204a; 1932 260–1; 1933a; 1936c; 1941b), he sometimes published insults that only a saint could entirely forgive (1922a 329; 1922b 86a; 1935f; 1937b 302a–318; 1939a 173a; 1941c 143; 1960 2, 4; SI 3, 76–7, 88, 91, 96, 100–2, 120, 141, 162–3). It is not evident that Fisher always struck the first blow in these quarrels (K. Pearson, presumably, in Soper et al. 1917 353; Bowley 1935 55–7; see also E. S. Pearson 1968; Eisenhart 1974), though their actual roots would be difficult if not impossible to trace (1922b 91; 1923b), nor did he always emerge the undisputed champion in bad manners<sup>5</sup> (K. Pearson 1936; Neyman 1951). On one occasion, Fisher (1954) struck out blindly against a young lady who had been anything but offensive or incompetent. His conclusion was that had the lady known what she was about she would have solved a certain problem in a certain fashion; he was right about that but failed to notice that she had solved it in just that fashion. Of course, Fisher was by no means without friends and admirers too. ‘Indeed, we are all his admirers. (Yet he has few articulate partisans in controversies on the foundations of statistical inference, the closest, perhaps, being Barnard (e.g. 1963) and Rao (e.g. 1961).)’

The main point for us in Fisher’s touchiness and involvement in quarrels is their impediment to communication (van Dantzig 1957; Yates 1962; Yates and Mather 1963; see also § 2.4). Those great statisticians who had the most to gain from understanding him, whether to some extent through their own tactlessness or otherwise, received the greatest psychological provocation to close their minds to him. Also, it is hard for a man so driven and so involved in polemic as Fisher was to recognize in himself and announce a frank change of opinion except when he is the first to see the need for it (1922a 326). For example, when Fisher says, ‘It has been proposed that . . .’ (SI 172), and then proceeds to smash that proposal to smithereens, would it occur to you that the proposer was Fisher himself (1935c 395)? ‘Yet specific, technical mistakes he can admit openly (1940b 423) and even gracefully (1930c 205), and he often mentions weaknesses of his earlier attempts which he later improved on (1922a 308a; 1922b 86a; 1925a 699a; 1930b 527a; 1930c 204a; SI 54, 56, 142).’

I am surely not alone in having suspected that some of Fisher’s major views were adopted simply to avoid agreeing with his opponents (Neyman 1961 148–9). One of the most valuable lessons of my rereading is the conclusion that while conflict may sometimes have somewhat distorted Fisher’s presentation of his views<sup>6</sup> (Yates 1962 1152), the views themselves display a steady and coherent development (Barnard 1963 164; Fisher 1920, 1922a, 1924a, 1925a, 1935b;

<sup>5</sup> There is good reason to think that Savage would have modified such bad manners of his own, but there is no way to eliminate them editorially without danger of distorting his intentions.

<sup>6</sup> Fisher says (SI 31) ‘[Chrystal’s] case as well as Venn’s illustrates the truth that the best causes tend to attract to their support the worst arguments, which seems to be equally true in the intellectual and in the moral sense.’

1926a, 1929d, DOE; Author's Notes in CMS on these and 1922b, 1922c). Ideas that I had consistently tuned out until the present reading are to be found in some of his earliest papers. (See individual topics below for references; see also 1928b, RW § 57.)

As in the works of other mathematicians, research for the fun of it is abundant and beautiful in Fisher's writings, though he usually apologizes for it' (1929c; 1942a 305a; 1953a; DOE § 35).

2.4. *Predecessors and contemporaries.* Fisher had a broad general culture and was well read in the statistical literature of his past and of his youth (GT; DOE xv; SI v; 1950b; 1958c; see also the rest of this subsection and the next). To begin with the oldest, the famous essay by Thomas Bayes (1763) seems to have been more stimulating to Fisher than to many who, like myself, are called Bayesians. Recognition of this was slow to come to me because of Fisher's rejection of Bayes' rule and other 'conventional' prior distributions (1921a 17; 1922a 311, 324–6; 1930b 528–31; 1932 257–9; 1934a 285–7; RW § 5; DOE § 3; SI Ch. 2), and because he certainly was not a Bayesian in any ordinary sense. His admiration for Bayes is to be inferred more from Fisher's attitude to inductive inference, which he sometimes explicitly links to Bayes (1930b 531; 1934a 285–6; 1936a 245–7; 1960 2–3; DOE § 3) and which will be discussed later, especially in Section 4.4, than by certain explicit words of praise (RW § 5; DOE § 3; SI 8–17; 1934a 285–6; 1936a 245–7. He urged the 1958 reprinting of Bayes (1763); see p. 295.).

Intellectually, Fisher was a grandson of Francis Galton, whom he greatly admired (1948 218; SI 1–2; yet DOE § 19 points out a serious error made by Galton), and a son of Karl Pearson, who was always before Fisher's eyes as an inspiration and a challenge (RW § 5; SI 2–4, 141; 1933b 893–4; see also § 2.1 and § 2.6 and Eisenhart 1974), so that Freud too might have called Pearson a father to Fisher.

Fisher always refers to "Student," William Sealy Gosset, with respect (1915 507–8; 1922a 315; 1922c 608; 1923a 655; 1924b 807–8; 1936a 252; 1938 16; 1939d; RW § 5; DOE 33; SI 4, 80–1; he disagrees politely in 1929f, less politely in 1936c), and their mutual admiration and enduring friendship is reflected in the collection of letters from "Student" to Fisher (Gosset 1962), which has the benefit of some annotation by Fisher.

Some of Fisher's important ideas about likelihood are anticipated in a long and obscure paper by Edgeworth (1908–9 esp. 506–7, 662, and most especially 82–5). When this was publicly pointed out to Fisher (Bowley 1935) he replied that there was nothing of value in Edgeworth's paper that was not in still older papers that Fisher was glad to acknowledge (1935b 77). Fisher seems to me to have underestimated the pertinent elements of Edgeworth's paper. I doubt that

<sup>7</sup> Nevertheless, the difficulty of documenting this assertion indicates that only a tiny fraction of Fisher's work is mathematical research for the fun of it.



Fisher ever read it all closely, either before or after the connection was pointed out, first because it is human to turn away from long and difficult papers [apparently] based on what one takes to be ridiculous premises—in this case, Bayes' rule and inverse probability—then later perhaps because it is hard to seek diligently for the unwelcome. Rao (1961 209–11) stresses that Fisher's contributions to the idea of maximum-likelihood estimation go far beyond those of all of his predecessors.<sup>8</sup>

In science, it is hostility rather than familiarity that breeds contempt, and all of Fisher's castigation of the Neyman–Pearson school (1934a 296; 1935c 393; 1935f; 1939a 173a, 180; 1945b 130; 1955; 1960; SI) shows that he never had sufficient respect for the work of that school to read it attentively, as will be brought out frequently in this essay. And members of that school in referring to Fisher were likely to read their own ideas impatiently into his lines. This too will be documented by implication during this essay. An interesting study on the breakdown in communication between the two sides might be based merely on the discussion following (Neyman 1935); and it might well begin with careful pursuit of the last complete paragraph on page 172 of that discussion. (See also references in § 2.3 and § 2.5.)

2.5. *Style.* Fisher's writing has a characteristic flavor, at once polished and awkward. It is not pleasant to my taste but is fascinating and can be inspiring. He has a tendency to be aphoristic and cryptic.<sup>9</sup> Sometimes things are plainly said—when you go back and check—but in such a way as to go unperceived or even undeciphered when first seen.

Mathematics is ruthlessly omitted from Fisher's didactic works, *Statistical Methods for Research Workers* and *The Design of Experiments*. In modern mathematical education there is great repugnance to transmitting a mathematical fact without its demonstration. The disciplinary value of this practice is clear, but, especially in the mathematical education of nonmathematicians, it can be abused. Many a high school boy knows more biology, chemistry, and physics, than a dozen men could demonstrate in a lifetime. Is it not then appropriate for him also to know more mathematics than he himself can demonstrate? Giving perhaps too affirmative a response (RW x, § 4), Fisher freely pours out mathematical facts in his didactic works without even a bow in the direction of demonstration. I have encountered relatively unmathematical scholars of intelligence and perseverance who are able to learn much from these books, but for most people, time out for some mathematical demonstrations seems indispensable to mastery (Hotelling 1951 45–6).

<sup>8</sup> Scraps of other drafts of this paragraph were nearby in Savage's manuscript. My contribution following this paper further describes and assesses Edgeworth's paper.

<sup>9</sup> Examples culled from Savage's notes include 1915; 1920 761; 1921 b 119, 122; 1922 c 598, 600; 1923 b § 2; 1924 a; 1924 b 807, 810; 1928 b; 1930 b; 1934 a 297; 1935 b 42, 47; RW § 57; see also van Dantzig (1957), Hotelling (1951 38, 45–6), and for his own view and excuses, Fisher (1922 b 86 a; 1926 a 511; see also § 2.3) and Mahalanobis (1938 265–6).



Fisher is not one to confine himself to technicalities. He sometimes sermonizes, 「on statistics proper and more broadly.」 Here are a few examples both good and bad; classification I leave as an exercise to the reader (DOE §§ 4, 16, 37, 66 par. 2; 1914; 1938; 1948; 1950b; 1958a; 1958b; 1960). He sometimes enters upon the history of ideas, usually not very well in my opinion (RW § 5; DOE §§ 2–3; SI Ch. 1–2; 1922a; 1930b; 1936a; 1936d; 1948; 1953c; 1958a). Like some other great men, he does not hesitate to castigate as childish the work of equally great men not only within but also outside areas of his competence (see § 2.3).

2.6. *Just what did Fisher do in statistics?* It would be more economical to list the few statistical topics in which he displayed no interest than those in which he did. To discuss his achievements with any thoroughness would require a long paper in itself. My object here is merely to say enough about them to set the stage for the discussion of Fisher's statistical point of view, which is the ultimate object of this paper. I shall often say, for simplicity of language, that Fisher was the innovator of various topics in statistics. I shall of course not do so where I definitely know that he had predecessors. At the same time, searching for predecessors is difficult if not all but impossible.

In the art of calculating explicit sampling distributions, Fisher led statistics out of its infancy (1915; 1921a; 1922c; 1924b), and he may never have been excelled in this skill (Neyman 1961 147). "Student" (1908) conjectured the distribution of "Student's"  $t$ ; Fisher proved it 「explicitly and in detail for one-sample  $t$  (1923a; see also 1939d) and extended it to  $t/(1 - r^2)^{1/2}$  (1915 518) and standardized coefficients in multiple regression (1922c).」 Fisher discovered the distribution of  $F$  (and<sup>1</sup> of its logarithm  $z$  (1921a; 1922c; 1924b relates the various problems). (Actually, 「the symbol」  $F$  was introduced by Snedecor (1934; 1937 § 10.5) in honor of Fisher, for which officiousness Fisher seems never to have forgiven him.<sup>10</sup>) A very sophisticated achievement of this sort was his early computation of the distribution of the sample correlation coefficient for correlated variables (1915), 「a late one his treatment of dispersion on a sphere (1953a).」

It stands to reason that he would not have investigated noncentral distributions, because their *raison d'être* is the power function, a concept on which Fisher turned his back (see § 4.7). So much the worse for reason; Fisher was the first to give formulas for the important noncentral distributions, chi-squared,  $t$ , and 'singly noncentral'  $F$  (1928a; 1931b).

<sup>10</sup> This attitude has not been found expressed in Fisher's writing, and Finney wrote, "I thought he always had considerable affection for Snedecor," yet Savage's statement seems from some comments received to reflect an oral tradition. Though Fisher tabulated  $F$  (ST) and used it in exposition (DOE § 23), he personally preferred  $z$  because its distribution varies more regularly and is more nearly normal, facilitating interpolation and tabulation (RW § 41, DOE § 23, ST 2). According to Eisenhart, he also considered its scale better for expressing departure from the null hypothesis, and its near normality helpful in combining independent analyses. His incidental remark (1924b 808) that  $z = \log s_1/s_2$  has mode  $\log \sigma_1/\sigma_2$  piqued Savage.

Fisher is the undisputed creator (Cochran 1976; Yates and Mather 1963 107–113; see also Hotelling 1951 42–3; Mahalanobis 1938 271; Neyman 1951 407; 1961 146–7; 1967 1458–9) of the modern field that statisticians call the design of experiments, both in the broad sense of keeping statistical considerations in mind in the planning of experiments and in the narrow sense of exploiting combinatorial patterns in the layout of experiments. 「His book *Design of Experiments* is full of wonderful ideas, many already clearly presented or present in (1926a).」 I shall mention quite a few of these, discussing one or two a little, but am in danger of leaving out several of your favorite ones by oversight. He preached effectively against the maxim of varying one factor at a time (1926a 511–2; DOE Ch. 6 esp. §§ 37, 39). He taught how to make many comparisons while basing each on relatively homogeneous material by means of partial replication<sup>11</sup> (1926a 513; 1929d 209–12; DOE Ch. 7–8). He taught what should be obvious but always demands a second thought from me: if an experiment is laid out to diminish the variance of comparisons, as by using matched pairs (which can be very useful), or by adopting a Knut Vik square (which presumably cannot be made very useful), the variance eliminated from the comparisons shows up in the estimate of this variance (unless care is taken to eliminate it) so that as actual precision is gained perceived precision can be lost (1926a 506–7; 1939d 7; DOE §§ 27, 33, 34). Randomized design, and perhaps even the notion of a rigorously random sample, seems to originate with Fisher (1925a 700–1; 1926a; RW § 48; DOE §§ 5, 6, 9, 10, 20 (which seems to claim priority), 22, 31; Cochran 1975; Neyman 1951), though this technique is so fundamental to modern statistics that to credit Fisher with it sounds like attributing the introduction of the wheel to Mr. So-and-So. Some combinatorial designs are so natural as to be unavoidable and still others, illustrated by the Knut Vik square, were familiar in agronomy when Fisher began work in the field, but he inaugurated the systematic study of combinatorial designs (1934c; DOE § 35), and introduced the main sophisticated categories of them (1926a 513; 1929d; DOE Ch. 7–8). The analysis of variance and the analysis of covariance are his terms and, admitting that everything has antecedents, presumably his inventions (1918 134, 424, 433; 1921b 110–11, 119–22; 1922c 600; 1923c 315–9; 1924b 810–13; RW Ch. 7–8; DOE see Index). Along with the analysis of variance goes the *F*-test—or *z*-test, as Fisher would prefer.

The design of experiments suggests an interesting digression illustrating how two great statisticians may move in entirely different worlds. Wald mentioned in his *Statistical Decision Functions* (1950) that since choosing what experiment to do is a decision, the theory of design is a topic under the general theory of

<sup>11</sup> Savage may have chosen these words to avoid more specific terms and keep the meaning general. As usual, the references are no clue as he did not supply them. “Confounding” or “partial confounding” would also make sense, especially if the previous sentence is taken to cover fractional factorial design. “Fractional replication” would make less sense and was invented by Finney, not Fisher.

that book, and this remark of Wald's was perhaps too ostentatiously repeated in publicity for the book, such as the jacket. To Fisher (1955 70), this claim was incomprehensible because Wald's book does not 'discuss elements of design such as replication, control, and randomization.'

Fisher must have been the first to have that very broad vision of regression—or the linear model—which is one of the most fertile insights of modern statistics (1921 b; 1922 c; RW §§ 25–9). In his day, the day of the desk calculator, it was natural to emphasize shortcuts in the calculations associated with regression<sup>12</sup> (RW §§ 26–9), so it is natural that Fisher does not greatly emphasize the study of residuals. Yet, he does sometimes study residuals, and I imagine that he is an originator here too (1921 b 122–6; 1922 a 322; 1924 c 108–10; RW § 28.1 par. 1).

Fisher with Tippett (1928 c) opened the field of the asymptotic distribution of extreme values (Gumbel 1958 3). Watson and Galton (1874) are commonly considered the fathers of the study of branching processes, but it was Fisher who brought generating functions to bear on this topic and thereby put it on the mathematical map<sup>13</sup> (1930 c).

Fisher invented and vigorously pursued  $k$ -statistics, unbiased estimators of cumulants (1929 a; 1930 a; 1931 a; 1937 a). This seems strange for a man who had no use for unbiasedness as a criterion in estimation (1915 520; 1935 b 42; SI 140), but I would not hasten to preclude that he had a reason perfectly consistent with his philosophy.<sup>14</sup> 'Fisher helped work out the maximum likelihood analysis of the probit model (1935 d; ST) along with Bliss (1935; 1938) who originated the name (1934). (The model itself is old (Fechner 1860) and was first used in biological assay by Gaddum (1933) and Bliss (1934; 1935); see Finney (1952 § 14).)'

From two early controversies, Fisher has emerged completely victorious. There used to be some confusion, and I infer, outright disagreement (described by Yates and Mather 1963 101), about how to count the degrees of freedom in a contingency table. 'Fisher's view (1922 b; 1923 b; 1924 a; RW §§ 5, 20) has prevailed over Karl Pearson's.' Likewise, it was difficult to convince Karl Pearson (1900; 1936) and presumably others, that moments might be inefficient

<sup>12</sup> This also explains Fisher's attention to grouping (1922 a 317–21, 359–63; 1937 b 306–14; RW see Index at "grouping" and "Sheppard's adjustment"; DOE end of § 21).

<sup>13</sup> One of Savage's  $3 \times 5$  cards indicates that he intended to check this, and he would surely have changed it if he had received, as I did, a letter from D. G. Kendall including the following information. The basic problem and all three cases of the criticality theorem were stated by Bienaymé (1845) in a paper rediscovered by Heyde and Seneta (1972). Watson did use generating functions, but made an error in the supercritical case. In genetics, Haldane (1927) has a careful, accurate statement of all three cases. For further history, see D. G. Kendall (1966, 1975).

<sup>14</sup> The reason he usually gives is that using population cumulants and their unbiased estimators greatly simplifies the equations which connect the moment functions of the sampling distributions of moment statistics with the moment functions of the population (1929 a 198 a, 203, 204, 237; 1930 a 15 a; 1937 a 4–5).

statistics for estimating the parameters of certain distributions, such as those in the Pearson family. 「Here too Fisher's view (1922a 321–2, 332–56; 1937b; RW §§ 13, 56) has prevailed.」

It has long been recognized that the name “Cramér–Rao Inequality” is historically unjust (Savage 1954 238). Though I have been interested in the history of this beautiful inequality, it had never even been suggested to me that Fisher had played a role in that history, yet I find . . .<sup>15</sup>

「Despite his preference for parametric methods (1925a 701; 1929f; 1936a 250; 1943 end of par. 3; DOE § 17 par. 2, § 23 last par.),」 Fisher was a pioneer in nonparametric statistics on the basis of his introduction of the sign test in lieu of a certain application of the *t*-test (RW § 24 Ex. 19) and still more his introduction of what he called exact tests to escape from the hypothesis of normality in many applications of *t* and *F*-tests (DOE § 21, 23 last par.; see also 1929f; 1936b). 「He also touched on runs (1926c) and on order statistics as fiducial limits for population percentiles (1939d 4–5; 1945b 131; SI 81–2) and tabulated normal scores (expected order statistics) and their sums of squares for use in two-sample tests and analysis of variance based on ranks (ST Introduction and Tables XX and XXI).」

Fisher seems to have been almost a nonparticipant in sequential analysis for no particular reason that I can discern. But even in this field he was an innovator (1922a § 12; 1941a; 1952a). 「He also developed, apparently independently (Barnard 1963 163), the idea of a minimax, randomized strategy while “solving” (as we would now say) the game of *le Her* (1934d).」

I have deliberately refrained from the discussion of certain original and far-reaching ideas of Fisher in the field of estimation and inference because discussion of these is part of my main object to which everything thus far has been an extended preamble.

**3. Basic technical concepts.** By reviewing certain technical concepts introduced by Fisher, or at any rate of central importance to him, we can obtain some ideas of his statistical outlook while remaining on relatively solid ground. This will prepare us both technically and mentally to study his more subtle and controversial positions.

Fisher coined the term “statistic” for 「a function of the data designed (usually, sometimes inexplicitly or unnecessarily) to estimate a parameter (1922a 313;

<sup>15</sup> Savage left this incomplete and may have changed his mind. Fisher (1925a) comes tantalizingly close, but Savage's notes on § 7 thereof say “Cramér–Rao would be a big help but not available.” His 3 × 5 cards on Cramér–Rao agree with this and also say, “I was confusing with Koopman–Pitman,” and “Late and little” with a reference to SI 145 ff. Perhaps he merely intended to mention the “rediscovery of a faint version” he mentions in par. 4 from the end of § 3 below. Another card says that Fisher (1934a) “scooped” Pitman–Koopman (on families of distributions admitting sufficient statistics). It is conceivable that Savage intended to refer to this instead, or even to use it to relate Fisher to the Cramér–Rao bound by way of the fact that the Pitman–Koopman families are the only regular ones achieving it. My contribution following this paper sketches another relationship.

1925a 701; RW § 11; see also CMS Index and 1922a esp. 309–10, 316–7, 329–31; RW §§ 1–3, 53–6). In its current meaning,<sup>1</sup> an arbitrary function of the data, not necessarily real-valued, the concept is extremely familiar today and recognized to be invaluable. The term for it is not necessarily the best imaginable, though it by now seems ineradicable.

Estimates are of central importance to Fisher, but I doubt that he attempted any precise definition of the concept. Perhaps we can safely say that an estimate is a statistic, especially a real-valued one, intended to estimate something. Sometimes in the writings of Fisher and other statisticians “estimate” is seen from the context to mean a sequence of estimates, one associated with each sample size (1922a; 1925a; 1935b; RW §§ 3, 53, 55, 56). When this is done, it is in a context in which the asymptotic behavior of the sequence of estimates has the stage. As Fisher came to feel, not only is this ellipsis excessive, but such asymptotic considerations lack a uniformity of control necessary for practical conclusions (SI 144–5). For example, if  $\bar{X}_n$  is asymptotically virtuous, then the sequence of estimates that are identically zero for  $n < 10^{1000}$  and equal to  $\bar{X}_n$  elsewhere has the same asymptotic virtues as  $\bar{X}_n$  and the same practical lack of virtue as using 0 for the estimate regardless of the data.

By “estimation,” Fisher normally means what is ordinarily called point estimation (see CMS or CP Index at “estimation”; RW § 2, Ch. 9; DOE § 66; SI Ch. 6). In particular, he does not refer to fiducial intervals as estimates (1935b 51). The term “point estimation” made Fisher nervous, because he associated it with estimation without regard for accuracy, which he regarded as ridiculous and seemed to believe that some people advocated (1935b 79; SI 141); this apprehension seems to me symptomatic of Fisher’s isolation from other modern theoretical statisticians<sup>16</sup> (§ 2.4).

The idea and terminology of a sufficient statistic or a set of sufficient statistics was introduced by Fisher in its current form (1920 768–9; 1922a 316–7, 331; 1925a 713–4; the latter two include factorization. See also Stigler 1973.). Whether a sufficient statistic deserved the term used to be controversial but Fisher has won hands down. I know of no disagreement that when an experiment admits a given statistic as sufficient then observation of that statistic is tantamount for all purposes to observation of all the data of the experiment.

Intimately associated with sufficient statistics is the concept of the likelihood of an experiment depending on a parameter, possibly multidimensional. The most fruitful, and for Fisher, the usual, definition of the likelihood associated with an observation is the probability or density of the observation as a function of the parameter, modulo a multiplicative constant; that is, the likelihood associated with an observation is the class of functions of the parameter proportional

<sup>16</sup> Nevertheless, his objection (SI 140–1) to common criteria of point and interval estimation because they lack invariance under single-valued transformations is justified in the sense that such criteria will draw distinctions among estimates which are equally informative in his sense (see below) and common sense.

to the probability of the observation given the parameter (1922a 310, 326–7, 331; 1925a 707; though not a probability, it may measure relative degree of rational belief: 1930b 532; 1932 259; 1934a 287; SI 66–73, 126–31; see also § 4.4). The likelihood of independent observations is the product of the likelihoods of each observation, and for this reason, it is often convenient to work with the logarithm of the likelihood (SI 71, 148).

The likelihood is a minimal sufficient statistic. That is, the likelihood of the data, regarded as a random object (in this case, a random function on parameter space), is sufficient for the data, and the likelihood can be recovered from any sufficient statistic. Fisher seems to have been the discoverer of this important fact, and he was very appreciative of it (1925a 699b; 1934a 287, 289, 294, 306; SI 49–50, 151).

‘Usually consistent estimation is defined’ to mean a sequence of estimates, one for each sample size, that converges in probability to the parameter being estimated. ‘Fisher gave a different definition, now usually called Fisher consistency. (1922a 309, 316; 1924a 444; SI 142, 144. An entry to current literature is Norden 1972–3.) He tended for some time to treat the usual definition as interchangeable with his (1924a 444; 1925a 702–3; 1935b 41–2; RW §§ 3, 53) but ultimately rejected it (SI 144).’ A Fisher-consistent estimate is mathematically a functional defined on distributions that coincides with the parameter to be estimated on the family of distributions governed by the parameter. Employed as an estimate, this functional is applied to the empirical distribution of a sample of  $n$  independent drawings from an unknown distribution of the family.<sup>17</sup> Though motivation can be seen for this definition, it has drawbacks. Many functions commonly thought to be excellent estimates are not consistent under this definition because they are not really functions of the empirical distribution. Certain favorite estimates of Fisher’s such as the  $k$ -statistics (for  $k > 1$ ) of which the ordinary estimate of the variance is the most important are not, strictly speaking, functions of the empirical distribution but of the empirical distribution and the sample size. Fisher does not seem to have mentioned this and would undoubtedly regard it as mathematical nitpicking.<sup>18</sup> On the other hand, I suspect that Fisher would have seen it as an advantage of this definition that it rendered “inconsistent” certain examples (J. L. Hodges, Jr., see Le Cam 1953 280) that had been invented

<sup>17</sup> Conceivably there are other ways of making Fisher’s definition precise, but to improve on this one would be hard. Von Mises (1964 Ch. 12 and references therein) used the term “statistical functions” and investigated their continuity, differentiability, laws of large numbers, and asymptotic distributions. Savage planned to “say somewhere why Fisher consistency tends to promise ordinary consistency.” The reason is, of course, that the empirical distribution converges (in probabilistic senses) to the true distribution, and hence, if a functional is appropriately smooth, its empirical value will converge to its true value.

<sup>18</sup> Here in the manuscript Savage had a note to himself saying, “Quote him about the Nile at least.” Fisher’s “problem of the Nile” (1936a 258, 244a; see also CP Index; SI 119, 163) is equivalent to a previously formulated and partially solved problem of similar regions according to Neyman (1961 148 footnote).



to show the mathematical inadequacy of Fisher's definition of efficiency. A particularly grave inadequacy of this definition of consistency is its inapplicability to any form of statistical data other than repeated samples from the same distribution. Thus, it is not directly applicable to observation of a stochastic process such as a normally distributed time series.<sup>19</sup>

Fisher (1922a 323–6) introduced the term “maximum likelihood estimate,” which is now so familiar to us all. He credits Gauss with early introduction of the method itself (1930b 531; 1936a 249; RW 21; perhaps his concern to distinguish the method from “inverse probability” explains why he does not mention Gauss in earlier papers discussing or using it: 1912; 1915 520; 1921a 4; 1922a; 1925a; M. G. Kendall (1961) traces the method to Bernoulli (1778) and Sheynin (1966 1004; 1971 § 3.2) to Lambert (1760). See also Edwards (1974).). There is general agreement today that maximum likelihood estimates are often excellent and that under certain circumstances they can act up. For example, the likelihood may be unbounded in the neighborhood of a point that is not an attractive estimate (Hill 1963). Or maximum likelihood may be very inappropriate when the number of parameters is not small compared with the number of observations (Neyman and Scott 1948).<sup>20</sup>

Fisher often confines his discussion, whether explicitly or implicitly, to applications of maximum likelihood in which the probabilities or densities are regular in the parameter, though he does sometimes also explore cases in which it is not.<sup>21</sup>

<sup>19</sup> It also excludes most admissible estimates for more-than-three-dimensional parameters, as Le Cam mentioned in correspondence. But see footnote 16.

<sup>20</sup> In the talk, Savage mentioned that Fisher might “cheat a little bit,” somewhat as follows. In multiparameter situations, if the ordinary maximum likelihood estimate of an individual parameter has a distribution depending only on that parameter, its likelihood can in turn be formed and maximized to produce a “second-stage” maximum likelihood estimate, typically different from the first. For  $\sigma^2$  in normal models, for instance, the first has denominator  $n$  while the second is the usual unbiased estimate and avoids the Neyman–Scott pathology. Fisher does not refer to or recommend the two-stage procedure in general (1922a 323–4; SI 152), but he occasionally practices something like it without noting the difference (1915 520 for  $\rho$ ; 1922c 600 for  $\sigma^2$ ). He may well be referring to this difference when he says, to clear up a point which “has been the cause of some confusion,” mentioning just these examples: “...it is not surprising that certain small corrections should appear, or not, according as the other parameters of the hypothetical surface are or are not deemed relevant.” (1922a 313. This does not explain how one would come initially to consider  $r$  or  $s$  in the latter case. See also 1912 157–9.) The difference for  $\rho$  seems to have confused K. Pearson and perhaps contributed to his break with Fisher (Eisenhart 1974 456) while that for  $\sigma^2$  puzzled “Student” (E. S. Pearson 1968 446–8). Edgeworth (1908 393–4) discusses the latter in the framework of inverse probability.

<sup>21</sup> In a long discussion of Pearson curves, Fisher (1922a 332–55) gives a little attention to the irregular effects of unknown endpoints on estimation of location, especially for Type III (gamma or chi square with location and scale parameters), but he ignores the repercussions for scale and shape. (In the irregular region he thus obtains smaller asymptotic variances for maximum likelihood estimates of scale and shape with location unknown than he would with location known.) He also discusses the double exponential density with unknown location (1925a 716–7; 1934a 297–300), where the irregularity causes no first-order problem but has an interesting second-order effect. I have not found any other irregular cases explored by Fisher.



Of course, the behavior of maximum likelihood estimation cannot be expected to be the same in regular and in irregular situations. (For further discussion and references, see below and § 4.2.)

The differential, or Fisher, information associated with a parameter is a function to which Fisher, quite properly, attached great importance (1922a 308a, 310, 329, 367; 1924a 445; 1925a 699a, 708–25; 1934a 298–306; 1935b 42–8; 1936a 249–50; 1938 17; RW §§ 55, 57.2, 57.3, 58; DOE § 60, Ch. 9; SI Ch. 6). In this, he was to some extent anticipated (Edgeworth 1908–9 esp. 502, 507–8, 662, 677–8, 82–5 and references he cites including Pearson and Filon 1898, also cited by Fisher), but Fisher's role in exploring this concept should not be underrated. The Fisher information plays a crucial role in what has been called the Cramér–Rao inequality, about which Fisher may have been completely ignorant, except for the rediscovery of a faint version of it (SI 145–6). There is, of course, a multivariate extension of the Fisher information, which Fisher knew and understood (1922a 332–6, 339; SI 152–5 with a spectacular mistake about matrices in eq. (169)).

Important properties of Fisher information to which Fisher called attention are that it is additive for independent observations (1925a 710; 1934a 298; 1935b 47; SI 149) and that it can only become smaller if the data are grouped or otherwise summarized (1925a 717; 1935b 44, 47; RW § 55; DOE § 73; SI 150–2). In particular, a statistic fails to lose information if and only if it is a sufficient statistic (1925a 699a, 717–8; 1935b 47; DOE § 74; SI 151).

Fisher recognized very early a formal connection between Fisher information and entropy (1935b 47). We know today that it is closely related to the entropy-like concept of Shannon–Wiener information as is explained, for example, in Kullback (1959 Ch. 1–2), Kullback and Leibler (1951), Savage (1954 § 15.6), M. G. Kendall (1973).

What is Fisher information good for? In the case of large samples, the reciprocal of the information tells with good accuracy the mean-squared deviation that can be achieved by good estimates, 'in particular, the maximum likelihood estimate,' over a reasonable range of the parameters, 'and in this connection Fisher often called it intrinsic accuracy (1922a; 1925a; 1935b; 1936a).' The exact mathematical facts are delicate; some references are Cramér (1946), Le Cam (1953), Hájek (1972). Fisher felt that his measure of information had a deeper and not purely asymptotic significance, and that is the first of the controversial points discussed in the next section.

#### 4. Points of controversy.

4.1. *The small-sample interpretation of information.* The large-sample interpretation of Fisher information has just been reviewed and is very well known. Despite the delicacy of certain theorems concerning it, it makes sense and is extremely valuable to the practicing statistician. A dramatic way to put the asymptotic conclusion, of which Fisher was very fond (1920 762; RW §§ 3, 49,

55; DOE § 60; see also 1922a 316; 1925a 704; 1938 17) is that statistics losing a fraction of the information lose that fraction of the work done to gather the data. This seems basically correct to me, and it is not so intimately bound up with variance as the measure of the inaccuracy of an estimate as might be thought from my description so far.

From my own point of view, the Fisher information is typically the reciprocal of the variance of a normal distribution which is a good approximation, in large experiments, of the posterior distribution of the parameter (under almost any prior). This asymptotic variance is an appropriate index of the width of the posterior distribution for almost any practical loss function.

But Fisher insisted that to lose information was tantamount to losing a corresponding number of observations even in small samples (1922a 338–9, 350–1; 1925a 669a, 709, 712, 714–22; 1934a 300; SI 152; see also below). At first, he seemed to expect this to speak for itself, but it met with doubt and even derision (Bowley 1935) so Fisher eventually developed what he called a two-story argument to justify his nomenclature and idea. If a large number of small experiments were done and the data from each replaced by some statistic of smaller information than the original experiment, then the many small experiments taken together would constitute a large experiment with  $n$  times the information of a component experiment and the  $n$  statistics taken together would constitute a large experiment with a fraction, say  $\alpha$ , of that information. This would indeed represent a waste of  $(1 - \alpha) \times n$  of the  $n$  small experiments (1935b 41, 46–7; SI 157). As an argument for saying that an estimate in a given small experiment wastes the fraction  $(1 - \alpha)$  of the total information in that experiment, I myself regard this more as a ploy than as a real move. (See also SI 159.) It does give a practical counsel in case one is determined to summarize each of a large number of small experiments in terms of one or few numbers, but this is not likely to be an applicable model to somebody who proposes to use the median, rather than the mean, of a sample of size 5 (of presumably guaranteed normality). But the argument does at least deserve to be known, and I, for one, found it a surprise on rereading Fisher.

4.2. *Properties of maximum likelihood estimation.* Are maximum likelihood estimates typically Fisher consistent? Fisher said they were (1935b 45–6; SI 148; see also 1922a 328 and references below on efficiency, which implicitly requires consistency) and with considerable justification. Consider first, as Fisher did, a multinomial experiment, that is a number of independent trials each of which ranges over the same finite set of possible outcomes with the same distribution depending on the unknown, possibly multivariate, parameter  $\theta$ . As Fisher emphasized, there is in principle no loss in generality in confining all discussions of statistics to experiments with finite numbers of outcomes and,

therefore, all discussions of repeated trials to multinomial experiments.<sup>22</sup> To my mind, this is an important lesson, though it can be carried too far, as I shall soon have occasion to illustrate. If  $f(i)$  is the frequency of occurrences of 'outcome'  $i$  in  $n$  trials of a multinomial experiment with probabilities  $P(i|\theta)$ , then the likelihood is expressible in terms of the empirical distribution; in fact, the log likelihood satisfies

$$\log L(\theta|f) \propto \sum \frac{f(i)}{n} \log P(i|\theta),$$

where  $\propto$  denotes proportionality. Thus the maximum likelihood estimate can fairly be said to be a function of the empirical distribution with the remark that it may be an incompletely defined function insofar as there may be distributions for which the maximum is not attained or is attained for more than one  $\theta$ . If in the logarithm of the likelihood  $f(i)/n$  is replaced by  $P(i|\theta_0)$  for some value  $\theta_0$  of the parameter, then, according to a well-known inequality made popular by modern information theory (see references in § 3, next-to-last par.) the expression attains an absolute maximum when  $\theta$  is  $\theta_0$ . Moreover, this maximum is unique if the parameter is (in a terminology not Fisher's) identified, that is, if  $P(i|\theta)$  is not the same function of  $i$  for two different values of  $\theta$ . Thus, with only the most reasonable qualification, maximum likelihood can be said to be Fisher consistent in multinomial experiments, and these have a good claim to being for all practical purposes the most general experiments to which the notion of Fisher consistency would be applicable.

Since the 'qualification' is a natural one in connection with 'consistency' and since every practical sequence of repeated trials can be adequately approximated by a multinomial experiment, this would perhaps seem to settle the whole matter for the practical statistician, but there is a snare in that argument, which I mention here to illustrate how the ubiquity of multinomial experiments can be misleading. The relatively easy theorems about the multinomial experiment depend on the possibility of choosing  $n$  so large that all of the  $f(i)$  will have high probability of being within a small fraction of a per cent of  $n \times P(i|\theta)$ . This can indeed be done, but for many natural multinomial experiments and, in particular, some natural multinomial approximations of typical continuous-variable experiments, the necessary value of  $n$  for this condition to be satisfied might well be, say,  $10^{30}$ , in which case the result gives no practical assurance whatsoever. There is at least hope that other lines of demonstration would be more reassuring.

With a little inventiveness, it seems possible to view the logarithm of a likelihood as a function of the empirical distribution which has a meaning for other

<sup>22</sup> I haven't found this point emphasized in Fisher's writing. Sometimes he mentions it (SI 50 and perhaps 143-4) or assumes without comment that considering a finite number of classes is sufficiently general (1925 a 700-1, 718; 1935 b 45; SI 142, 145). Eisenhart reports that in a 1951 conversation, Fisher said that he made the point in 1922a, but all Eisenhart found was the second sentence of § 12. The only reference I found in Savage's notes is 1924 a; all I see there is that classes are used, but they are required for chi square (and similarly in 1928 b).

distributions (possibly taking on the value  $-\infty$ ) which would make Fisher consistency of the maximum likelihood estimate true in very great generality. Whether there is real use, or only a certain satisfaction of the sense of magic, in knowing that maximum likelihood estimation can be said to be Fisher consistent, I cannot say.

Fisher of course expected maximum likelihood estimates to be consistent in the sense of convergence in probability also (see references on consistency in § 3 above). Certain kinds of exceptions he would have regarded as mathematical caviling. Indeed, this might be the case for any exceptions thus far known, 'conceivably even Bahadur's (1958).<sup>7</sup> Knowing Fisher, I am not surprised at my inability to find discussions of counterexamples, nor would I be surprised if some discussion were turned up somewhere. A mathematically satisfactory account of consistency in probability of maximum likelihood estimates has had a painful evolution and may not yet be complete. (See for example Wald 1949, Perlman 1972. Norden 1972–3 surveys various properties of maximum likelihood estimates, with a few idiosyncratic, neo-Fisherian touches.)

In smooth and civilized repeated trials, and many other kinds of large experiments, maximum likelihood estimation is not only consistent but efficient, that is, the distribution of the maximum likelihood estimate is approximately normal around  $\theta$  with the variance of the approximating distribution being the reciprocal of the Fisher information. (This does not mean that the variance of the estimate itself is that small or even finite (Savage 1954 242–3). But that is not the sort of distinction that I would expect Fisher to make or even countenance.) The tendency of maximum likelihood estimates to be efficient was appreciated by Edgeworth (1908–9) and later by Fisher (1922a 331–2, 367; 1922c 598; 1924a 445; 1925a 707, 710–11; 1932 260; 1935b 44–6; RW §§ 3, 46, 55, 58; SI 148). Neither succeeded in demonstrating the phenomenon with much generality from a modern mathematical point of view, though Fisher went inestimably further than Edgeworth. (See also § 2.4 and the end of § 3.)

Fisher asserted/conjectured that the maximum likelihood estimate alone among Fisher-consistent estimates has any chance of being a sufficient statistic (1922a 331; 1925a 714; 1932 259), and at first that it is always sufficient (1922a 323, 330, 367), later that it is sufficient whenever there exists a sufficient estimate (1922a 308a; 1935b 53, 82) or statistic (1922a 331; 1925a 714, 718; 1932 259; RW § 3; SI 151; he may mean a Fisher-consistent estimate in every case).<sup>7</sup> For me, it is not the business of an estimate to be sufficient, so I regard the question as somewhat distorted. Academic though the situation is, I have sought, and offer below, a counterexample.

Fisher also conjectured that no other Fisher-consistent estimate (or perhaps even more general kind of estimate) loses so little information as the maximum likelihood estimate (1925a 699a, 720–1, 723; 1932 260; 1935b 53; 1936a 249, 250–1, 256; SI 157). This point too is academic but curiosity provoking. The conjecture is false as stated, though there may be some way to reconstrue it that

makes it true. A counterexample based on repeated tosses of a pair of not necessarily fair coins, one of which rolls out of sight with known positive probability, is spelled out below. 「It avoids the triviality of counterexamples involving restricting the natural range of a parameter, such as a normal distribution with mean known to be positive, or a normal, random-effects model.」 Readers uninterested in the details may skip to § 4.3.

Let

$$P(i, j | \theta) = \frac{1}{8}(1 + i\theta)(1 + j\theta)$$

for  $i = -1, 0, 1$ , for  $j = -1, 1$ , and for  $-1 \leq \theta \leq 1$ . This is, for each  $\theta$ , a probability measure on the six values of the pair  $(i, j)$  and thus defines a multinomial process.

$$\begin{aligned} \frac{\partial}{\partial \theta} \log L(\theta | f) &= \sum_{i,j} f(i, j) \left( \frac{i}{1 + i\theta} + \frac{j}{1 + j\theta} \right) \\ &= \left( \frac{g(1)}{1 + \theta} - \frac{g(-1)}{1 - \theta} \right) \end{aligned}$$

for  $-1 < \theta < 1$ , 「where  $g(1)$  is the total frequency of 1's and  $g(-1)$  of -1's. (Thus  $g(k) = 2f(k, k) + f(k, -k) + f(0, k) + f(-k, k)$  for  $k = -1, 1$ .)」 The maximum likelihood estimate of  $\theta$  is therefore

$$\hat{\theta}(f) = \frac{g(1) - g(-1)}{g(1) + g(-1)},$$

which is defined for all possible  $f$  with  $\sum f(i, j) = n$  and  $n > 0$ , and it lies in the range of  $\theta$ ,  $[-1, 1]$ . (The sole purpose of the  $j$ -coin is to keep the denominator of  $\hat{\theta}(f)$  positive.)

The maximum likelihood estimate  $\hat{\theta}$  is not sufficient (for 'any  $n$ '), because  $\hat{\theta}$  can 'be  $\pm 1$ ' without determining 'both  $g(1)$  and  $g(-1)$ ', which constitute a minimal sufficient statistic. Therefore  $\hat{\theta}$  loses some Fisher information.

For each probability distribution  $P$  defined on the domain of  $f$ , namely  $i = -1, 0, 1$  and  $j = -1, 1$ , let

$$\begin{aligned} S(P) &= E(\pi^{i+2j} | P) \\ &= \sum_{i,j} P(i, j) \pi^{i+2j}, \end{aligned}$$

where  $\pi$  is a transcendental number larger than 1. Then  $S(f/n)$  is a sufficient statistic for  $\theta$ , since  $f/n$  can be reconstructed from  $S(f/n)$  because of the transcendence of  $\pi$ .

Let

$$\begin{aligned} Q(\theta) &= S(P | \theta) \\ &= E(\pi^{i+2j} | \theta) \\ &= E(\pi^i | \theta) E(\pi^{2j} | \theta) \\ &= \frac{1}{8} (\sum_i \pi^i + \theta \sum_i i \pi^i) (\sum_j \pi^{2j} + \theta \sum_j j \pi^{2j}). \end{aligned}$$

On  $[-1, 1]$ ,  $Q$  is the product of two positive and strictly increasing functions of  $\theta$ , so the inverse function  $Q^{-1}$  is well defined there.

Finally,  $Q^{-1}(S(f/n))$  is a Fisher-consistent (and also in-probability consistent) estimate of  $\theta$ , which unlike the maximum likelihood estimate  $\hat{\theta}$  is sufficient and therefore loses no Fisher information.

The estimate  $Q^{-1}(S(f/n))$  does not have mean squared error as small as that of  $\hat{\theta}$ , at least for large  $n$ . Can a cleverer example achieve even that?<sup>23</sup>

4.3. *What is probability?* Fisher, as everybody knows, was a frequentist, yet I—who profess to interest in such things—was somewhat taken aback in my rereading to find how vehemently he denies that probability is a limiting frequency in an indefinite sequence of repeated ‘‘actual’’ trials, which is the position that frequentists ordinarily take<sup>24</sup> (1935 b 81; 1958 a; 1960 5–7; SI 109–10; see also 1922 a 312–3; 1925 a 700–1; DOE §§ 3, 6; SI 14, 32–3, 44, 114–6).

For Fisher, a probability is the fraction of a set, having no distinguishable subsets, that satisfies a given condition (SI 32–3, 55–7, 109–10; 1958 a; 1960 5–6; see also 1922 a 312–3; 1925 a 700–1; 1935 b 78; 1945 b 129; 1955 75; 1961 a 3, 7). For example, in showing the cards of a well-shuffled deck one at a time, exactly one-quarter of the showings will result in hearts—there can be no question about that. And there are no distinguishable subsets; that means that no particular subset of showings, such as the first thirteen or either of the alternate sets of twenty-six, can be expected to be richer in hearts than others. This is a notion that not everyone will find clear and acceptable, but let us at least allow Fisher to describe it in his own terms:

For the validity of probability statements about the real world there are I believe only *three* necessary and sufficient requirements. (i) As Kolmogoroff rightly insisted now

<sup>23</sup> A page of Savage’s text is omitted here. It begins, ‘‘According to Fisher, the maximum likelihood estimate is the only Fisher-consistent estimate that is determined by a linear equation in the frequencies (19 ; ). There seems to be no truth in this at all...’’ Fisher states and shows this for estimates which are efficient, not merely consistent; that is, in current terminology, he shows that the only efficient  $M$ -estimator is the maximum likelihood estimator (1928 b 97–8; 1935 b 45–6; SI 148. Edgeworth did much the same: see my contribution following this paper.). The nearest I have found to Savage’s version is one sentence (SI 157) where Fisher doesn’t clearly impose either restriction and has just mentioned consistency but is concerned with distinguishing among ‘‘the different possible Efficient estimates’’ (SI 156). In view of this context and the earlier references, I think Savage’s version is a misreading of Fisher which would have been caught before publication.

<sup>24</sup> ‘‘Actual’’ has been inserted, at the risk of misrepresenting frequentists, because in early papers Fisher defines probability as a proportion in an ‘‘infinite hypothetical population’’ of what seem to be repeated trials under the original conditions, where ‘‘the word infinite is to be taken in its proper mathematical sense as denoting the limiting conditions approached by increasing a finite number indefinitely.’’ (1925 a 700; see also 1922 a 312.) Later he says, ‘‘An imagined process of sampling... may be used to illustrate.... Rather unsatisfactory attempts have been made to define the probability by reference to the supposed limit of such a random sampling process.... The clarity of the subject has suffered from attempts to conceive of the ‘limit’ of some physical process to be repeated indefinitely in time, instead of the ordinary mathematical limit of an expression of which some element is to be made increasingly great.’’ (SI 110.)

many years ago every statement of mathematical probability implies a mathematically well-defined Reference Set of possibilities, which must be measurable at least so far that members of the Set, comprising a known fraction  $P$  of the whole, possess some characteristic which is absent from the remainder. (ii) The subject, or particular entity about which the probability statement is asserted, must be a member of this Set. (iii) No sub-set may be recognizable having a fraction possessing the characteristic differing from the fraction  $P$  of the whole. (1960 5)

In a statement of probability the predicand, which may be conceived as an object, as an event, or as a proposition, is asserted to be one of a set of a number, however large, of like entities of which a known proportion,  $P$ , have some relevant characteristic, not possessed by the remainder. It is further asserted that no subset of the entire set, having a different proportion, can be recognized. (SI 109)

The reference set, as Fisher calls it, may well be infinite, where an infinite set is conceived of by Fisher as a sort of limit of finite sets (1925a 700–1; SI 110). Such a notion is hard to formulate mathematically, and indeed Fisher's concept of probability remained very unclear, which must have contributed to his isolation from many other statistical theorists. (See also §§ 2.3–2.5 and references there.)

4.4. *Statistical inference.* An important current of thinking in modern statistics, established by Neyman (1938; see also 1934 623), takes the point of view that a logic of the uncertain such as is suggested by the phrase “statistical inference” is illusory, but Fisher deplored that direction (SI 7, 100; see also 28–30, 34 and RW § 2 last par.; DOE § 2 first par.; 1930b 531; 1934a 287; 1935b 39–40; 1960 2–4), and always sought fervently to establish a “genuine” theory of statistical inference. According to Fisher, this valid goal was clearly and admirably established by Thomas Bayes, whom Fisher greatly admired, ‘not least for perceiving the weakness in “Bayes’ rule,” that is, resort to conventional prior distributions, such as the uniform distribution for an unknown frequency, which he severely condemns (for references, see par. 1 of § 2.4).’ Fisher’s approach might well be called “inductive logic” (a term he sometimes used: 1955 69; 1960 2; see also 1935b title and 39–41; 1958a 261) to contrast it with the phrase “inductive behavior” by which Neyman sought to suggest that the dilemmas of inductive inference could be solved by casting the problems of induction in the framework of Robinson Crusoe economics (see § 4.5).

Genuine statistical inference must, according to Fisher, issue with a definite, clear, and objective conclusion about the uncertain (DOE § 2; 1935b 39–40, 54; 1936a; 1960 2–4; see also 1921a 3; 1922a 311; 1930b 531; 1934a 285–7;



1939a 175; 1955 77; SI 2, 37, 106–10). For example, it might be expressed in terms of the probabilities of events, as is surely appropriate in those cases where there is an undisputed prior distribution (1922a 324; 1930b 530; 1932 257–8; 1934a 286; 1957 205; 1958a 272; SI 11, 17, 35, 111). This probability might be fiducial probability. I shall try to say more about fiducial probability later (§ 4.6). For the moment, I am content to explain that Fisher at first tried to introduce a different kind of probability applicable in some cases in which ordinary probability was not (1930b 532–5), but later came to hold that these probabilities were ordinary probabilities, serving the purpose of posterior probabilities in a Bayesian calculation, though arrived at by extra-Bayesian means (SI 51, 56; 1960 5; see also § 4.6).

But the conclusion of a statistical inference might be something other than a probability (1930b 532; 1934a 284a; 1955 76–7; 1960 4; SI 35, 131–6, Ch. 3.). For example, it might be a likelihood (1912 160; 1921a 24; 1922a 326–7; 1925a 707; 1930b 532; 1932; 1935b 40–1, 53, 82; SI 66–73, 126–31; see also §§ 3 and 4.8). Because likelihoods are intimately associated with probabilities, it has been suggested that the whole concept is superfluous (van Dantzig 1957 190). Yet, a likelihood function of a parameter, which might rightly be called a set of likelihood ratios, is evidently not a probability distribution for the parameter. Thus we can see why one who, like Fisher, believes that a likelihood function constitutes a statistical inference, would see here an example of a statistical inference that is not expressed in terms of probabilities, more exactly, in terms of a probability distribution of the unknown parameters.

Fisher often refers to exact tests (see § 4.7), so tests would seem to be for him a form of exact non-Bayesian inference issuing in tail areas which are neither likelihoods nor parameter distributions.

If nothing else can be said about induction, there will be general agreement that induction differs from deduction in this. Anything that can be deduced from part of the information at hand can be deduced from all of it, but in induction account must be taken of all of the data. Fisher is very fond of this point (1935b 54; 1935c 392–3; 1936a 254–5; 1937c 370; 1945b 129; 1955 75, 76; 1958a 268, 272; 1960 4, 10; SI 55, 109) though he lapses a bit on at least one occasion.<sup>25</sup>

Fisher seems to think the ignoring of pertinent information an essential feature of Neyman–Pearson statistics (1935c 393; 1955 76; SI 101; 1960 4, 7; see also below). There is at least one rather clear case in point. It has been suggested by Bartlett and followed up by Scheffé that to test whether two sets of  $n$  numbers have the same mean, though possibly different variances, the elements of the

<sup>25</sup> Anywhere that Fisher countenances the use of less than fully informative or efficient statistics could be considered an example, but presumably Savage had something more specific in mind. Unfortunately, the only possible reference I found in his notes (1939a 175) doesn't seem to be it. Eisenhart thinks Savage is probably referring to Fisher's use of order statistics as fiducial limits for population percentiles (§ 2.6 above.).

two sets might be "randomly" paired and then the  $n$  differences be subjected to a  $t$ -test. "[Bartlett never advocated this test in practice, and Scheffé, if he did, does not now. See Bartlett 1965 and Scheffé 1970 for later views and earlier references.]" "What," Fisher once asked me orally, "would the proponent of such a test say if it turned out to be significant at the 99% point, but if his assistant later discovered that hardly any pairing other than the one accidentally chosen resulted in so extreme a significance level?" (See also 1937c 375; RW § 24.1 Ex. 21; SI 96–9.) Choosing one among the many possible pairings at random and ignoring the results of those not examined but available for examination does constitute a sort of exclusion of pertinent evidence. However, there seems to me to be a very similar fault in all those applications of randomization that Fisher so vigorously advocated. Whenever we choose a design or a sample at random, we ordinarily are able to see what design or what sample we have chosen, and it is not fully appropriate to analyze the data as though we lacked this information, though Fisher in effect recommends that.

It should in fairness be mentioned that, when randomization leads to a bad-looking experiment or sample, "Fisher said that" the experimenter should, with discretion and judgment, put the sample aside and draw another. He speculated, for example, that a more complicated theory might make it possible to choose Latin squares at random from among the acceptable Latin squares. A few references harmonious with this point of view are<sup>26</sup> (Grundy and Healy 1950; Youden 1956–72; for further discussion and references, see Savage 1962 33–4, 88–9).

4.5. *Inductive behavior.* As already indicated, Fisher thought an economic approach to statistics no substitute for statistical inference (1955 69–70, 73–5, 77; SI 1, 4–5, 75–8, 99–103; see also par. 1 of § 4.4). In later works, he hinted that it might have its mundane applications for the slaves of Wall Street and the Kremlin (1955 70) but not for a free scientist in search of truth.

<sup>26</sup> I first thought Savage intended to refer to Fisher here, but I have found nothing, and Yates and Mather (1963 112) say that Fisher never faced up to the problem. The nearest hint I have found in Savage's notes is the comment "Chooses a square at random but not quite," referring to (1926a 510) where he has marked the following passage: "Consequently, the term Latin Square should only be applied to a process of randomization by which one is selected at random out of the total number of Latin Squares possible; or at least, to specify the agricultural requirement more strictly, out of a number of Latin Squares in the aggregate, of which every pair of plots, not in the same row or column, belongs equally frequently to the same treatment." The context gives no suggestion that what Fisher has in mind here is bad randomizations, and restricting randomization to one standard square or one transformation set seems more likely to me. My impression of his writing generally is of a hard-line view. Indeed, in the same paper, the last paragraph of the previous section and the last sentence of the following page (both also noted by Savage), distinctly suggest this, though conceivably Fisher's only concern in these passages is his frequent one of systematic arrangements (see § 2.6). In a 1952 conversation, however, when Savage asked Fisher what he would do if he happened to draw a Knut Vik square at random, Fisher "said he thought he would draw again and that, ideally, a theory explicitly excluding regular squares should be developed" (Savage 1962 88). Perhaps Fisher took a softer line privately than he felt appropriate for public exposition.

It is important that the scientific worker introduces no cost functions for faulty decisions, as it is reasonable and often necessary to do with an Acceptance Procedure. To do so would imply that the purposes to which new knowledge was to be put were known and capable of evaluation. If, however, scientific findings are communicated for the enlightenment of other free minds, they may be put sooner or later to the service of a number of purposes, of which we can know nothing. The contribution to the improvement of Natural Knowledge, which research may accomplish, is disseminated in the hope and faith that, as more becomes known, or more surely known, a great variety of purposes by a great variety of men, and groups of men, will be facilitated. No one, happily, is in a position to censor these in advance. As workers in Science we aim, in fact, at methods of inference which shall be equally convincing to all freely reasoning minds, entirely independently of any intentions that might be furthered by utilizing the knowledge inferred. (SI 102–3)

... I am casting no contempt on acceptance procedures, and I am thankful, whenever I travel by air, that the high level of precision and reliability required can really be achieved by such means. But the logical differences between such an operation and the work of scientific discovery by physical or biological experimentation seem to me so wide that the analogy between them is not helpful, and the identification of the two sorts of operation is decidedly misleading. (1955 69–70)

For my own part, it seems likely that any principles so general as to apply to all that could be called business should apply to scientific activity too—at least as a rough model. However, in the view of a personalistic Bayesian like me, the contrast between behavior and inference is less vivid than in other views. For in this view, all uncertainties are measured by means of probabilities, and these probabilities, together with utilities, guide economic behavior, but the probability of an event for a person (in this theory) does not depend on the economic opportunities of the person.

Fisher's hostility to inductive behavior seems somewhat inconsistent with his other views. For he is very much interested in diminishing the costs of experiments (1929d esp. 206; DOE generally, and specifically §§ 9 end, 12 end, 25, 31, 37 end, 55 par. 5, 60, 71 end), and writes about the cash objectives of agronomy as something important and not apart from his other statistical interests (1952a 186–7). Also, almost in the same breath with criticism of 'purely mathematical

statisticians whose theories refer to economics or decision functions,<sup>1</sup> Fisher warns that if his methods are ignored and their methods used a lot of guided missiles and other valuable things will come to grief (1958a 274).

4.6. *The fiducial argument.* The expressions “fiducial probability” and “fiducial argument” are Fisher’s. Nobody knows just what they mean, because Fisher (SI 56, 172) repudiated his most explicit, but definitely faulty, definition and ultimately replaced it only with a few examples (cited below; for definitions, see 1930b 532–5; 1935c 391–5; SI 51–60, 117–9, 169–73; Barnard 1963 166; also Fisher, DOE §§ 62 end, 63 end, and examples; 1936a 252–3, 255; 1945b 130–2; 1955 76–7; 1958a 271–3; 1960 5, 9–10). There still seem to be serious attempts to make something systematic out of Fisher’s fiducial ideas (Dempster 1968; Fraser 1961, 1968; see also references in Savage 1954, 2nd ed. (1972) 262).

In a word, Fisher hopes by means of some process—the fiducial argument—to arrive at the equivalent of posterior distributions in a Bayesian argument without the introduction of prior distributions (see reference to definitions above and, especially explicitly, 1939a177; 1945b 132; RW § 23 last par.; SI 51, 56, 80, 120, 125). The kind of attempt, its futility, and Fisher’s dogged blindness about it all seem to me very clear in the following passage:

The objection has been raised that since any statement of probability to be objective must be verifiable as a prediction of frequency, the calculations set out above cannot lead to a true probability statement referring to a particular value of  $T$  [observed], for the data do not provide the means of calculating this. This seems to assume that no valid probability statement can be made except by the use of Bayes’ theorem. However, the aggregate of cases of which the particular experimental case is one, for which the relative frequency of satisfying the inequality statement is known to be  $P$ , and to which all values of  $T$  are admissible, could certainly be sampled indefinitely to demonstrate the correct frequency. In the absence of a prior distribution of population values there is no meaning to be attached to the demand for calculating the results of random sampling among *populations*, and it is just this absence which completes the demonstration that samples giving a particular value  $T$ , arising from a particular but unknown value of  $\theta$ , do not constitute a distinguishable sub-aggregate to which a different probability should be assigned. Probabilities obtained by a fiducial argument are objectively verifiable in exactly the same sense as are the probabilities assigned in games of chance. (SI 58–9)

Notable features of this passage are Fisher’s expectation of and belief in purely

logical paths that lead to objective inference. Notwithstanding Bayes' own failure, which Fisher so clearly recognizes, Fisher holds this goal to be absolutely necessary to the advance of science (see § 4.4). This is why he regards its abandonment as "rather like an acknowledgment of bankruptcy" (1955 75).

All in all, there are a modest handful of fiducial distributions explicitly adduced by Fisher [for a single normal sample, those of  $\mu$ ,  $\sigma$  individually and jointly (1933a 346–7; 1935c 391–2, 395; 1936a 251–2; 1939d 4; 1941c 142–5; 1945b 130; 1955 75; RW § 23; DOE §§ 62, 63; SI 80, 119–20),  $\mu + \alpha\sigma$  with  $\alpha$  given (1941c 146; SI 121–3), and the mean and standard deviation of a future sample (1935c 393–5; SI 115–17); for a bivariate normal sample, of  $\rho$  (1930b 533–4; 1955 76), all parameters (SI 169–73), and  $\mu_1/\mu_2$  (1954); for matched pairs, of  $\mu_1/\mu_2$  (RW xiv, § 26.2; DOE § 62.1); for two independent normal samples, of  $\mu_1$ ,  $\sigma_1$ , and  $\sigma_2$  when  $\mu_1 = \mu_2$  (1961a 5–8; 1961b),  $\mu_1 - \mu_2$  when  $\sigma_1 \neq \sigma_2$  (1935c 396–7; 1937c; 1939a; 1941c; 1945b 132; 1961a 3–4; SI 94–6), and  $\mu_1$ ,  $\mu_2$  when  $n_1 = n_2 = \sigma_1 = \sigma_2 = 1$  and  $\mu_1$ ,  $\mu_2$  are unrestricted or lie on a line or half-line (1955 77; SI 132–4); for other normal models, of regression coefficients (SI 84; DOE § 64, 201–2) and intersections (RW § 26.2) and of components of variance (1935c 397–8); for the exponential model, of the parameter (SI 52–4) and the total of a future sample (SI 113–4); in general, of the parameter of any monotonic, one-parameter model admitting a single sufficient statistic (1930b 532–4; 1934a 292–3; SI 69–70) and of the fractions of an arbitrary continuous population exceeded by the order statistics (1939d 4–5; 1945b 131; SI 81–2); see also below).<sup>7</sup> The most conspicuous and important of these coincide with Bayesian posterior distributions from standard priors. For example, the fiducial distributions of the mean and variance of a normal distribution, after repeated measurements, are those adduced from a uniform prior distribution on the mean and the logarithm of the variance. Fisher emphasizes that not all of his fiducial distributions coincide with posterior distributions (1930b 534–5; 1933a 347–8; SI 55).

Closely associated with, but I think of more lasting importance than, fiducial probability is what Fisher called ancillary statistics. An ancillary statistic is one the distribution of which does not depend on a parameter about which an inference is to be made. [Most of Fisher's later discussions seem to require independence of all unknown parameters (SI 134–5, 160–1, 166, 168; 1960 10; 1961a 6; in 1936a 256 this requirement is also mentioned, but seems not to be part of the definition). A definition of particular interest below appears in 1935b 48: ancillary statistics "by themselves supply no information on the point at issue." Other, especially earlier, discussions give no explicit requirement and concern recovery of information lost in estimation, including the "true weight" or "apparent precision" (1925a 699a, 724; 1932 260; 1934a 300, 307; 1935b 48; 1936a 256; 1955 72; SI 158–9).<sup>7</sup>

For example, in some study to determine the regression of a length as a function of temperature, it might well be that lengths are distributed normally about

a linear function of temperature regardless of what chance or purposive processes may lead to the temperatures. In this case, the whole set of temperatures is ancillary. Though the distribution of an ancillary statistic, such as the temperatures in the example, does not depend on the parameter of interest, it may be, as the example clearly shows, vital in making inferences about the parameter (SI 84–5).

Fisher believes, and most of us with him, that if the statistic is ancillary, inference can be made from the conditional distribution of the data, given the parameters of interest and the ancillary statistic (1934a 300–1, 305; 1935b 48; 1936a 257; 1955 71–2; SI 84–5, 161, 167). That, for example, is how everyone ordinarily studies the regression coefficient even if the sample is drawn from a bivariate normal distribution. Many of us, in learning statistics, have caught ourselves saying that the regression coefficient in this case has a certain  $t$ -distribution. Then there comes a realization that this  $t$ -distribution depends on the variance of the “temperature.” Observing that, one is perhaps tempted to try to calculate the ‘marginal’ distribution of the regression coefficient, but ‘the conclusion’ of statisticians of all persuasions has seemed to be that the conditional distribution of the regression coefficient given the values of the temperatures is ‘appropriate’ for inference.

As a personalistic Bayesian, I see the situation thus. From the temperatures, I would ordinarily have no new information about the ‘population’ regression coefficient. The reason is that given the population ‘mean and’ variance of the temperatures, the distribution of the actual temperatures, including their sample variance, is literally independent of the ‘population’ regression coefficient, and ordinarily, the ‘population mean and’ variance of the temperatures will in my personal judgment be irrelevant to the regression coefficient. This latter condition, which is not automatic, is sometimes overlooked by non-Bayesians, because they do not have easy ways to express it. These conditions imply that when I am given the temperatures, my opinion about the regression coefficient is unaffected, so that I can, without any real difference, study the effect of the rest of the data on my opinion about the regression coefficient, regarding the temperatures as given.

An important moral is that if the ‘population mean and’ variance of the temperatures were actually given, then the temperatures would indeed be ancillary in the ‘most restrictive’ technical sense, but when we regard the temperatures as ancillary to the regression coefficient in a bivariate normal distribution, our argument depends in part on how we choose to organize the parameters of the distribution and is not so objective as it may seem.<sup>27</sup>

<sup>27</sup> This relates to the ambiguity of “does not depend” in the first (less restrictive) definition of “ancillary” four paragraphs back. Specifically, we are interested in the parameters of the regression of  $x_2$  on  $x_1$ , and in the residual variance, and if we choose  $\mu_1$  and  $\sigma_1$  as the nuisance parameters, then the distribution of  $x_1$  does not depend on the parameters of interest, i.e., it depends  
(continued on next page)

In what I infer is Fisher's last major discussion of fiducial probability, he combines the notion of ancillary statistics with his very first notion of fiducial probability in a clever way (SI 159–69; see also 1935 b 51–4). By regarding certain statistics as ancillary for a single parameter  $\mu$ , it may be that the whole experiment can be regarded as a 'measure of  $\mu$ ' by means of a single sufficient statistic  $x$  describable by a family of cumulative distributions  $F(x|\mu)$ . Now, if this function happens to be monotone in  $\mu$  (as it is in  $x$ ) and if, for each  $x$ , it ranges over the interval from 0 to 1, then it can be regarded as a probability distribution in  $\mu$ . This is the fiducial distribution and is of course closely akin to the example quoted above. In such a case, at least if no ancillary statistics are invoked, this process is so similar to a familiar one for obtaining confidence intervals that it has given rise to the misconception that confidence intervals and fiducial intervals are two words for the same thing. This has been the source of much misunderstanding (as pointed out in 1935 c; 1937 c; 1939 c; 1955 74–7; 1960 4; SI 60, 64–5; Bartlett 1939; Neyman 1941; 1956 291–3; 1961 149).

†Fisher applies this process explicitly to two independent normal observations having means  $\mu_1, \mu_2$  restricted to lie on a circle, with the observed distance from the center ancillary (SI 134–5); to the bivariate density  $\exp\{-\theta x - y/\theta\}$ , with  $U = (\sum x_i)^{\frac{1}{2}}(\sum y_i)^{\frac{1}{2}}$  ancillary (SI 163–9; the generalization  $\theta^{s+1} \exp\{-\theta x - \theta^s y\}$  for any given  $s$  is also mentioned); to arbitrary location or location-and-scale models, with the "configuration" or "complexion" of the sample ancillary (SI 160–2; see also 1934 a 284 a, 300–6; 1936 a 256–7); and to a dilution-series model, with another kind of configuration ancillary (1935 b 51–4, 78). He also derives fiducial inequalities for a  $2 \times 2$  table taking the margins as ancillary (1935 b 38 a, 50–1, 79; but compare 1955 77; SI 62–6, 70).‡

The most famous of Fisher's applications of fiducial probability is to the Behrens–Fisher problem. The problem is not necessarily of great practical

---

only on the nuisance parameters. This is not the case, however, if we choose  $\mu_1$  and  $\rho$  as the nuisance parameters. The former allows  $\bar{x}_1, s_1$  to be ancillary, the latter  $r$  (perhaps—see below). The meaning of "does not depend on certain variables (or parameters)" depends on how the remaining variables (parameters) are chosen. A Bayesian definition would include the condition that the distribution of ancillary statistics depends only on nuisance parameters which are also judged to be a priori independent of the parameters of interest. This helps prevent a contradictory multiplicity of ancillary statistics. Any definition should also require, I believe, that the distribution of the observations given the ancillary statistics depend only on the parameters of interest, i.e., that the ancillary statistics be sufficient for the nuisance parameters when the parameters of interest are known (Kendall and Stuart 1961 217). This holds trivially for the more restrictive definition and may be implicit in Savage's discussion. If it is required, then neither  $\bar{x}_1$  nor  $s_1$  nor  $r$  is individually ancillary, but  $\bar{x}_1$  and  $s_1$  still are jointly. Though this requirement obviously reduces multiplicity, it by no means resolves all problems of conditional inference. As Savage (1962 20) says, "For a striking, if academic, example, suppose  $x$  and  $y$  are normal about 0 with variance 1 and correlation  $\rho$ . Then  $x$  and  $y$  are each by themselves irrelevant to  $\rho$ , and each is an ancillary statistic for the total observation  $(x, y)$  by any criterion known to me." See also Birnbaum (1962 esp. 279), Cox (1958 esp. 359–63), Dawid (1975), Hájek (1967 esp. 150–4).



importance (RW § 24.1 par. before Ex. 20; SI 93; see also 1935c 395; 1939a 180; 1941c 149), but it vividly illustrates a difference in conclusion between Fisher and frequentists of the Neyman–Pearson school. The problem is to estimate the difference between the means of two normal distributions of not necessarily equal variance. The fiducial distribution of each mean is that of the sample mean plus a  $t$ -like variable times the sample standard deviation and these two population means are fiducially independent. Therefore, their difference is fiducially distributed like the difference between the two sample means plus a linear combination of two independent  $t$ -like variables (references eight paragraphs above). The fiducial intervals thus adduced are known not to be confidence intervals,<sup>28</sup> and they command no respect from adherents of the Neyman–Pearson school (Bartlett 1965 § 3; Scheffé 1970 footnote 4). For Jeffreys, who accepts uniform priors for the unknown means and for the logarithms of the variances, what Fisher calls the fiducial distribution of the difference of the two means is simply its posterior distribution. Indeed, Jeffreys claims to have preceded ‘Fisher’ in the discovery of this answer, and apparently with justice.<sup>29</sup>

4.7. *Hypothesis testing.* Hypothesis testing was extremely important to Fisher, and his ideas about it do not coincide with those that are now most widely known through the influence of the Neyman–Pearson school. Let him speak for himself

<sup>28</sup> This means (presumably) that there exist  $n_1, n_2, \sigma_1/\sigma_2$ , and  $\alpha$  for which the coverage probability of the fiducial intervals is less than  $1 - \alpha$ . Fisher writes as if this couldn’t happen (1939a 173a; 1960 8; SI 96). To demonstrate that it can, there is apparently only one published example, one of three coverage probabilities calculated by E. L. Scott and given by Neyman (1941, table near end of § 4; according to Scott, the headings should be corrected to read  $n = 7$ ,  $n' = 13$ , and  $\rho^2$ ). This example is, however, in contradiction with a table of Wang (1971 Table 5) for the same  $\alpha$  (two-tailed .05) and degrees of freedom (6 and 12), where the error rate is given for variance ratios 1/32 to 32 by factors of 2, is everywhere  $< .05$ , and varies far too smoothly to be compatible with Scott’s value (.066 at a variance ratio of 10. According to D. L. Wallace, who drew my attention to her paper, Wang’s values are within .0002 except at a variance ratio of 32, where the correct error rate is .0491, not .0499.). Furthermore, calculations by Geoffrey Robinson (1976) show one-tailed error rates less than  $\alpha$  for  $\alpha = .1, .05, .025, .01, .005, .001, .0005, .0001$ ;  $n_1, n_2 = 2(1)8(2)14, 18, 24, 32, 50, 100, \infty$ ; and  $\sigma_1^2 n_2 / \sigma_2^2 n_1$  or  $\sigma_2^2 n_1 / \sigma_1^2 n_2 = 1, 1.5, 2, 3, 5, 10, 30, 100, 1000$ , which he considers sufficient “to infer with reasonable certainty” that even the one-tailed procedure is conservative. Mehta and Srinivasan (1970) and Lee and Gurland (1975) found some (one-tailed) error rates above  $\alpha$  at variance ratios near 0 and  $\infty$  for a second-order asymptotic approximation to the fiducial procedure. Elsewhere their values are appreciably below  $\alpha$ . The fiducial procedure has error rate exactly  $\alpha$  at ratios 0 and  $\infty$ . This suggests that adjusting their values to apply to the exact fiducial procedure would give error rates everywhere below  $\alpha$  in the cases they calculate also.

<sup>29</sup> Savage left space for references after “claims” and “justice,” but no such claim has been found, and it is absent from two notable discussions of the problem by Jeffreys (1939 § 5.42; 1940). In the one-sample problem Jeffreys (1931 69; 1932) preceded and indeed instigated Fisher’s (1933a) paper.

Incidentally, D. J. Finney notes that Fisher himself insisted on referring to the Behrens distribution and test and disliked “Behrens–Fisher,” and especially “Fisher–Behrens,” even asking Finney to correct a misuse of his. Fisher’s writing is consistent with this, and he states (SI 94) that his (1935c) paper “confirmed and somewhat extended Behrens’ theory.”

on several important points:

... it is convenient to draw the line at about the level at which we can say: "Either there is something in the treatment, or a coincidence has occurred such as does not occur more than once in twenty trials." ...

If one in twenty does not seem high enough odds, we may, if we prefer it, draw the line at one in fifty (the 2 per cent point), or one in a hundred (the 1 per cent point). Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and ignore entirely all results which fail to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment *rarely fails* to give this level of significance. (1926a 504)

Our examination of the possible results of the experiment has therefore led us to a statistical test of significance, by which these results are divided into two classes with opposed interpretations.... The two classes of results which are distinguished by our test of significance are, on the one hand, those which show a significant discrepancy from a certain hypothesis; ... and on the other hand, results which show no significant discrepancy from this hypothesis.... In relation to any experiment we may speak of this hypothesis as the "null hypothesis," and it should be noted that the null hypothesis is never proved or established, but is possibly disproved, in the course of experimentation. Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis.

... It is evident that the null hypothesis must be exact, that is free from vagueness and ambiguity, because it must supply the basis of the "problem of distribution," of which the test of significance is the solution. A null hypothesis may, indeed, contain arbitrary elements, and in more complicated cases often does so: as, for example, if it should assert that the death-rates of two groups of animals are equal, without specifying what these death-rates actually are. In such cases it is evidently the equality rather than any particular values of the death-rates that the experiment is designed to test, and possibly to disprove.

In cases involving statistical "estimation" these ideas may be extended to the simultaneous consideration of a series of hypothetical possibilities. The notion of an error of the so-called "second kind," due to accepting the null hypothesis

“when it is false” may then be given a meaning in reference to the quantity to be estimated. It has no meaning with respect to simple tests of significance, in which the only available expectations are those which flow from the null hypothesis being true. (DOE § 8)

The attempts that have been made to explain the cogency of tests of significance in scientific research, by reference to hypothetical frequencies of possible statements, based on them, being right or wrong, thus seem to miss the essential nature of such tests. A man who “rejects” a hypothesis provisionally, as a matter of habitual practice, when the significance is at the 1 % level or higher, will certainly be mistaken in not more than 1 % of such decisions. For when the hypothesis is correct he will be mistaken in just 1 % of these cases, and when it is incorrect he will never be mistaken in rejection. This inequality statement can therefore be made. However, the calculation is absurdly academic, for in fact no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas. Further, the calculation is based solely on a hypothesis, which, in the light of the evidence, is often not believed to be true at all, so that the actual probability of erroneous decision, supposing such a phrase to have any meaning, may be much less than the frequency specifying the level of significance. To a practical man, also, who rejects a hypothesis, it is, of course, a matter of indifference with what probability he might be led to accept the hypothesis falsely, for in his case he is not accepting it. (SI 41–2)

Though recognizable as a psychological condition of reluctance, or resistance to the acceptance of a proposition, the feeling induced by a test of significance has an objective basis in that the probability statement on which it is based is a fact communicable to, and verifiable by, other rational minds. The level of significance in such cases fulfils the conditions of a measure of the rational grounds for the disbelief it engenders. It is more primitive, or elemental than, and does not justify, any exact probability statement about the proposition. (SI 43)

There are many unusual features here. The importance of the power function and error of the second kind is vehemently denied (further references below),

as is the possibility of testing any hypothesis other than a sharp one, that is, one that postulates a specific value for a parameter [or a function of parameters] (but see also below and SI 46, 89–92). Apparently there have been statisticians who recommended actually picking a level before an experiment and then rejecting or not according as that level was obtained. I do not have the impression that any professional statisticians make that recommendation today, though it is still often heard among those who are supposed to be served by statistics, but Fisher's strong rejection of the notion is noteworthy (SI 43; but compare 1926a 504, DOE §§ 7, 61). Though the importance, or even the existence, of power functions is [sometimes] denied, Fisher says that some tests are 'more sensitive' than others, and I cannot help suspecting that that comes to very much the same thing as thinking about the power function. (DOE §§ 8, 11, 12, 61; SI 21, 42, 47–8; see also RW § 2 footnote, § 18 Ex. 5, § 24 Ex. 19; 1926a 504; 1934a 294–6. Fisher argues that failure to reject the null hypothesis does not establish it and hence is not an "error": 1935e; 1955 73; see also DOE §§ 8, 61.)

The logic of "something unusual" is very puzzling, because of course in almost any experiment, whatever happens will have astronomically small probability under any hypothesis. If, for example, we flipped a coin 100 times to investigate whether the coin is fair, all sequences have the extremely small probability of  $2^{-100}$  if the coin is fair, so something unusual is bound to happen. Once when I asked Fisher about this point in a small group, he said, "Savage, you can see the wool you are trying to pull over our eyes. What makes you think we can't see it too?" At any rate, the doctrine of "something unusual" does not work if taken very literally, and this, of course, is why Fisher had recourse to tail areas, grouping outcomes as more or less antagonistic to a given null hypothesis (DOE §§ 7, 8; see also 1926a 504; 1936a 251–2; and references below).

For Fisher, it was very important that tests be "exact" (DOE § 17 par. 2; 1936a 251, 252; 1939a 174; 1939d 2, 5–6; see also § 4.4). For this, it would be enough that they be exact given appropriate ancillaries, as I have illustrated in the discussion of regression. Often, "exact" seems to mean having a given size in the Neyman–Pearson sense (1960 8; DOE §§ 7, 8, 61; SI 37–9, 87, 96; see also references below). This, however, does not serve to explain Fisher's use of the Behrens–Fisher distribution in testing whether two means are equal in the presence of possibly unequal variances (1935c 397; 1939a; 1945b; 1960 8; 1961a; SI 94–6; he argues that "repeated sampling from the same population" is misleading and other reference sets are appropriate, without fully explaining his reference set in the Behrens–Fisher problem, even in 1961a, and in general without fully reflecting the fact that the expectations of conditional levels are unconditional levels: 1939a 173a–b; 1945b 130, 132; 1945c; 1955 70–2 but compare 73; 1960 6–7; 1961a; SI 39–44, 75–103).

4.8. *The likelihood principle.* The likelihood principle is a doctrine that seems in retrospect very appropriate to Fisher's outlook, though it does not seem to

have been plainly stated by him until his last book (SI 70–1, 72–3, 136; see also 1932 259; 1934a 287; 1935b 40–1; 1936a 249). Indeed, the first formal statement of the likelihood principle known to me is not by Fisher but by Barnard<sup>30</sup> (1947, 1949). The principle is still controversial, but I believe that it will come to be generally accepted. That the likelihood is a minimal sufficient statistic is an objective technical fact (see § 3). That such a statistic is as useful as the whole data for any statistical purpose is never really denied. (A seeming denial sometimes arises when critics point out that in practice specific statistical models can never be wholly trusted so that a statistic sufficient on the hypothesis of a given model is not sufficient under the wider hypothesis that that model may not actually obtain.) Thus, no one doubts that the likelihood function together with a statement of the distribution of this function for each value of the unknown parameter would constitute all that is relevant about an experiment bearing on the parameter. The likelihood principle goes farther, however: it says that the likelihood function for the datum that happens to occur is alone an adequate description of an experiment without any statement of the probability that this or another likelihood function would arise under various values of the parameter.

In a certain passage (SI 72–3), Fisher seems pretty forthrightly to advocate the likelihood principle. It could be argued that he means to apply it only to “statistical evidence of types too weak to supply true probability statements” (SI 70).

Fisher does sometimes depart from the likelihood principle. For example, ‘tail probabilities and hence significance tests do so.’<sup>31</sup> More disturbingly, a Poisson process admits a fiducial inference if the number of arrivals is fixed (SI 52–4) but not if the total time is fixed, despite identical likelihoods. This and ‘another, similar example’ are given in (Anscombe 1957).

According to Fisher, when other devices, such as Bayes’ theorem and the fiducial argument, are not available, the likelihood constitutes in itself an exact statistical inference (see § 4.4). Late in his work (SI 71), Fisher suggests a sort of test which is not a tail area test but consists simply in reporting the ratio of the maximum of the likelihood under the null hypothesis to its maximum under the alternate hypothesis.<sup>32</sup>

<sup>30</sup> Barnard writes that Fisher’s statement (RW § 2 last par., already in 1st ed.) is as formal. Savage’s “random note” on Fisher (1936a 249) says, “Full likelihood principle clear here if not earlier,” but a  $3 \times 5$  card questions this. In earlier works Savage says that it was “first put forward,” and “emphasized to statisticians” by Barnard (1947) and Fisher (SI), and gives further discussion (1962 17–20) and references (1954 2nd (1972) ed. iv), including one to Birnbaum (1962), who certainly gives a formal statement.

<sup>31</sup> Savage noted that the likelihood principle is “in effect denied” by Fisher (1922a 314 last par. 2nd sentence).

<sup>32</sup> Savage’s manuscript unfortunately breaks off here. He intended to add a section on errors and inconsistencies, but drafted no further text. What notes he left suggest that he intended to  
(continued on next page)

## REFERENCES

- ANSCOMBE, F. J. (1957). Dependence of the fiducial argument on the sampling rule. *Biometrika* **44** 464-469.
- BAHADUR, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhyā* **20** 207-210.
- BARNARD, G. A. (1947). Review of *Sequential Analysis* by Abraham Wald. *J. Amer. Statist. Assoc.* **42** 658-664.
- BARNARD, G. A. (1949). Statistical inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **11** 115-149.
- BARNARD, G. A. (1963). Fisher's contributions to mathematical statistics. *J. Roy. Statist. Soc. Ser. A* **126** 162-166.
- BARTLETT, M. S. (1936). The information available in small samples. *Proc. Cambridge Philos. Soc.* **32** 560-566.
- BARTLETT, M. S. (1939). Complete simultaneous fiducial distributions. *Ann. Math. Statist.* **10** 129-138.
- BARTLETT, M. S. (1965). R. A. Fisher and the last fifty years of statistical methodology. *J. Amer. Statist. Assoc.* **60** 395-409.
- BARTLETT, M. S. (1968). Fisher, R. A. *Internat. Encycl. Social Sci.* **5** 485-491. Crowell, Collier, & Macmillan, New York.
- BAYES, THOMAS (1763). An essay towards solving a problem in the doctrine of chances, with Richard Price's Foreword and Discussion. *Philos. Trans. Roy. Soc. London Ser. A* **53** 370-418. Reprinted with a biographical note and some editing by G. A. Barnard in *Biometrika* **45** 293-315 (1958) and in Pearson and Kendall (1970).
- BERNOULLI, DANIEL (1778). The most probable choice between several discrepant observations and the formation therefrom of the most likely induction (with comment by L. Euler) (in Latin). *Acta Acad. Petrop.* 3-33. English translation by C. G. Allen with an introductory note by M. G. Kendall in *Biometrika* **48** 1-18 (1961), reprinted in Pearson and Kendall (1970).
- BIENAYMÉ, I. J. (1845). De la loi de multiplication et de la durée des familles. *Soc. Philomathique de Paris, Bull.: Extraits des procès verbaux des séances, Ser. 5* 37-39.
- BIRNBAUM, ALLAN (1962). On the foundations of statistical inference (with discussion). *J. Amer. Statist. Assoc.* **57** 269-326.
- BLISS, C. I. (1934). The method of probits. *Science* **79** 38-39. Correction *ibid.* 409-410.
- BLISS, C. I. (1935). The calculation of the dosage-mortality curve. *Ann. Appl. Biol.* **22** 134-167. The comparison of dosage-mortality data. *Ibid.* 307-333.
- BLISS, C. I. (1938). The determination of dosage-mortality curves from small numbers. *Quart. J. Pharmacy & Pharmacology* **11** 192-216.
- BOWLEY, A. L. (1928). *F. Y. Edgeworth's Contributions to Mathematical Statistics*. Roy. Statist. Soc., London.
- BOWLEY, A. L. (1935). Discussion of Fisher (1935 b). *J. Roy. Statist. Soc.* **98** 55-57.
- COCHRAN, WILLIAM G. (1976). Early development of techniques in comparative experiments. *On the History of Statistics and Probability* (D. B. Owen, ed.). Marcel Dekker, New York.
- COX, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist.* **29** 357-372.
- CRAMÉR, HARALD (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

add little beyond this. He ended his Fisher lecture, which was much more informal and somewhat more conjectural, and of course presented without documentation, with an invitation to the audience:

I do hope that you won't let a week go by without reading a little bit of Fisher. I'm sure that some of the things I've told you are incredible. Check up on some of those damn lies—it'll do you good. And you'll enjoy it!

- DAWID, A. P. (1975). On the concepts of sufficiency and ancillarity in the presence of nuisance parameters. *J. Roy. Statist. Soc. Ser. B* **37** 248–258.
- DEMPSTER, A. P. (1968). A generalization of Bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **30** 205–247.
- EDGEWORTH, F. Y. (1908–9). On the probable errors of frequency-constants. *J. Roy. Statist. Soc.* **71** 381–397, 499–512, 651–678. Addendum *ibid.* **72** 81–90.
- EDWARDS, A. W. F. (1974). The history of likelihood. *Internat. Statist. Rev.* **49** 9–15.
- EISENHART, CHURCHILL (1974). Karl Pearson. *Dictionary of Scientific Biography* **10** 447–473. Scribner's, New York.
- FECHNER, G. T. (1860). *Elemente der Psychophysik* (2 vols.). Breitkopf und Härtel, Leipzig.
- FIELLER, E. C., CREASY, MONICA A. and DAVID, S. T. (1954). Symposium on interval estimation (with discussion). *J. Roy. Statist. Soc. Ser. B* **16** 175–222.
- FINNEY, D. J. (1952). *Probit Analysis: A Statistical Treatment of the Sigmoid Response Curve*. Chapman and Hall, London.
- FISHER, RONALD AYLMEY. See separate list below.
- FRASER, D. A. S. (1961). The fiducial method and invariance. *Biometrika* **48** 261–280.
- FRASER, D. A. S. (1968). *The Structure of Inference*. Wiley, New York.
- GADDUM, J. H. (1933). Reports on biological standards. III. Methods of biological assay depending on a quantal response. *Med. Res. Council Special Report Series* no. 183.
- GOSSET, W. S. ("Student," q.v.) (1962). *Letters from W. S. Gosset to R. A. Fisher: 1915–1936*. 4 vols., plus a vol. of summaries by R. A. Fisher, with a foreword by L. McMullen. 2nd ed. (1970) in 1 vol. Privately circulated.
- GRUNDY, P. M. and HEALY, M. J. R. (1950). Restricted randomization and quasi-Latin squares. *J. Roy. Statist. Soc. Ser. B* **12** 286–291.
- GUMBEL, E. J. (1958). *Statistics of Extremes*. Columbia Univ. Press.
- HÁJEK, JAROSLAV (1967). On basic concepts of statistics. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 139–162. Univ. of California Press.
- HÁJEK, JAROSLAV (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. Univ. of California Press.
- HALDANE, J. B. S. (1927). A mathematical theory of natural and artificial selection, V. *Proc. Cambridge Philos. Soc.* **23** 838–844.
- HEYDE, C. C. and SENETA, E. (1972). Studies in the history of probability and statistics XXXI. The simple branching process, a turning point test and a fundamental inequality: A historical note on I. J. Bienaymé. *Biometrika* **59** 680–683.
- HILL, BRUCE M. (1963). The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *J. Amer. Statist. Assoc.* **58** 72–84.
- HOTELLING, HAROLD (1951). The impact of R. A. Fisher on statistics. *J. Amer. Statist. Assoc.* **46** 35–46.
- JEFFREYS, HAROLD (1931). *Scientific Inference*. Cambridge Univ. Press. (4th ed. 1973.)
- JEFFREYS, HAROLD (1932). On the theory of errors and least squares. *Proc. Roy. Soc. Ser. A* **138** 48–55.
- JEFFREYS, HAROLD (1939). *Theory of Probability*. Oxford: Clarendon Press. (4th ed. 1967.)
- JEFFREYS, HAROLD (1940). Note on the Behrens–Fisher formula. *Ann. Eugenics* **10** 48–51.
- KENDALL, D. G. (1966). Branching processes since 1873. *J. London Math. Soc.* **41** 385–406.
- KENDALL, DAVID G. (1975). The genealogy of genealogy branching processes before (and after) 1873. *Bull. London Math. Soc.* **7** No. 3 225–253.
- KENDALL, M. G. (1961). Studies in the history of probability and statistics. XI. Daniel Bernoulli on maximum likelihood. *Biometrika* **48** 1–2. Reprinted in Pearson and Kendall (1970). (See also Bernoulli (1778).)
- KENDALL, M. G. (1963). Ronald Aylmer Fisher, 1890–1962. *Biometrika* **50** 1–15. Reprinted in Pearson and Kendall (1970).
- KENDALL, M. G. (1973). Entropy, probability and information. *Internat. Statist. Rev.* **41** 59–68.
- KENDALL, M. G. and STUART, A. (1961). *The Advanced Theory of Statistics: Vol. 2, Inference and Relationship*. Charles Griffin, London.



- KULLBACK, SOLOMON (1959). *Information Theory and Statistics*. Wiley, New York.
- KULLBACK, SOLOMON and LEIBLER, R. A. (1951). On information and sufficiency. *Ann. Math. Statist.* **22** 79–86.
- LAMBERT, J. H. (1760). *Photometria, Sive di Mensura et Gradibus Luminis Colorum et Umbrae*. Augustae Vindelicorum, Augsburg, 1760, art. 271–306.
- LE CAM, LUCIEN (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. *Univ. of California Publ. in Statist.* **1** 277–329.
- LEE, AUSTIN F. S. and GURLAND, JOHN (1975). Size and power of test for equality of means of two normal populations with unequal variances. *J. Amer. Statist. Assoc.* **70** 933–941.
- MAHALANOBIS, P. C. (1938). Professor Ronald Aylmer Fisher. *Sankhyā* **4** 265–272. Reprinted in Fisher (CMS) and in *Biometrics* **20** 238–250 (1964).
- MEHTA, J. S. and SRINIVASAN, R. (1970). On the Behrens–Fisher problem. *Biometrika* **57** 649–655.
- NEYMAN, JERZY (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection (with discussion). *J. Roy. Statist. Soc.* **97** 558–625. Reprinted in Neyman (1967').
- NEYMAN, JERZY (1935). With K. Iwazskiewicz and St. Kołodziejczyk. Statistical problems in agricultural experimentation (with discussion). *J. Roy. Statist. Soc. Suppl.* **2** 107–180. Reprinted in Neyman (1967').
- NEYMAN, JERZY (1938). L'estimation statistique traitée comme un problème classique de probabilité. *Actualités Scientifiques et Industrielles* **739** 25–57. Reprinted in Neyman (1967').
- NEYMAN, JERZY (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32** 128–150. Reprinted in Neyman (1967').
- NEYMAN, JERZY (1951). Review of *Contributions to Mathematical Statistics* by R. A. Fisher. *Scientific Monthly* **62** 406–408.
- NEYMAN, JERZY (1956). Note on an article by Sir Ronald Fisher. *J. Roy. Statist. Soc. Ser. B* **18** 288–294.
- NEYMAN, JERZY (1961). Silver jubilee of my dispute with Fisher. *J. Operations Res. Soc. Japan* **3** 145–154.
- NEYMAN, JERZY (1967). R. A. Fisher (1890–1962): An appreciation. *Science* **156** 1456–1460.
- NEYMAN, JERZY (1967'). *A Selection of Early Statistical Papers of J. Neyman*. Univ. of California Press.
- NEYMAN, JERZY and SCOTT, ELIZABETH L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16** 1–32.
- NORDEN, R. H. (1972–3). A survey of maximum likelihood estimation. *Internat. Stat. Rev.* **40** 329–354, **41** 39–58.
- PEARSON, E. S. (1968). Studies in the history of probability and statistics. XX. Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments. *Biometrika* **55** 445–457. Reprinted in Pearson and Kendall (1970).
- PEARSON, E. S. (1974). Memories of the impact of Fisher's work in the 1920's. *Internat. Statist. Rev.* **42** 5–8.
- PEARSON, E. S. and KENDALL, M. G., eds. (1970). *Studies in the History of Statistics and Probability*. Charles Griffin, London.
- PEARSON, KARL (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **50** 157–175.
- PEARSON, KARL (1936). Method of moments and method of maximum likelihood. *Biometrika* **28** 34–59.
- PEARSON, KARL and FILON, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philos. Trans. Roy. Soc. London Ser. A* **191** 229–311.

- PERLMAN, MICHAEL D. (1972). On the strong consistency of approximate maximum likelihood estimators. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 263-281. Univ. of California Press.
- PÓLYA, G. (1937). Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen. *Acta Math.* **68** 145-253.
- PÓLYA, G. (1941). Heuristic reasoning and the theory of probability. *Amer. Math. Monthly* **48** 450-465.
- PÓLYA, G. (1962-5). *Mathematical Discovery*. (2 vols.) Wiley, New York.
- RAO, C. RADHAKRISHNA (1961). Apparent anomalies and irregularities in maximum likelihood estimation. *Bull. Inst. Internat. Statist.* **38** No. 4 439-453. Discussion *ibid.* No. 1 193-214. Reprinted in *Sankhyā Ser. A* **24** 73-101 (1962).
- ROBINSON, G. K. (1976). Properties of Student's *t* and of the Behrens-Fisher solution to the two means problem. *Ann. Statist.* **4**. To appear.
- SAVAGE, LEONARD J. (1954). *The Foundations of Statistics*. Wiley, New York. (2nd ed., Dover, New York, 1972.)
- SAVAGE, LEONARD J., et al. (1962). *The Foundations of Statistical Inference. A Discussion*. Methuen, London.
- SCHÉFFÉ, HENRY (1970). Practical solutions of the Behrens-Fisher problem. *J. Amer. Statist. Assoc.* **65** 1501-1508.
- SHEYNIN, O. B. (1966). Origin of the theory of errors. *Nature* **211** 1003-1004.
- SHEYNIN, O. B. (1971). J. H. Lambert's work on probability. *Arch. History Exact Sci.* **7** 244-256.
- SNEDECOR, GEORGE W. (1934). *Analysis of Variance and Covariance*. Collegiate Press, Ames, Iowa.
- SNEDECOR, GEORGE W. (1937). *Statistical Methods*. Iowa State College Press.
- SOPER, H. E., YOUNG, A. W., CAVE, B. M., LEE, A. and PEARSON, K. (1917). On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher. A cooperative study. *Biometrika* **11** 328-413.
- STIGLER, STEPHEN M. (1973). Studies in the history of probability and statistics. XXXII. Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* **60** 439-445.
- "STUDENT" (W. S. GOSSET, q.v.) (1908). The probable error of a mean. *Biometrika* **6** 1-25.
- VAN DANTZIG, D. (1957). Statistical Priesthood II: Sir Ronald on scientific inference. *Statistica Neerlandica* **11** 185-200.
- VON MISES, RICHARD (1964). *Mathematical Theory of Probability and Statistics*. (Ed. and complemented by Hilda Geiringer.) Academic Press, New York and London.
- WALD, ABRAHAM (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595-601.
- WALD, ABRAHAM (1950). *Statistical Decision Functions*. Wiley, New York.
- WANG, Y. Y. (1971). Probabilities of Type I errors of Welch tests for the Behrens-Fisher problem. *J. Amer. Statist. Assoc.* **66** 605-608.
- WATSON, H. W. and GALTON, FRANCIS (1874). On the probability of the extinction of families. *J. Anthropol. Inst. Great Britain and Ireland* **4** 138-144.
- YATES, F. (1962). Sir Ronald Fisher, F. R. S. *Nature* **195** 1151-1152.
- YATES, F. and MATHER, K. (1963). Ronald Aylmer Fisher. *Biographical Mem. Fellows Roy. Soc. London* **9** 91-120. Reprinted in Fisher (CP vol. 1).
- YOUNDEN, W. J. (1956-72). Randomization and experimentation. *Technometrics* **14** 13-22 (1972). Abstract in *Ann. Math. Statist.* **27** 1185-1186 (1956).

### CITED WORKS BY RONALD AYLMEY FISHER

- RW *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, 1925; 14th ed., 1970, reprinted by Hafner, New York, 1973. References are to 10th ed., 1948 reprinting, but section and example numbers are unchanged in successive editions.

- GT *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press, 1930; 2nd ed., Dover, New York, 1958.
- DOE *The Design of Experiments*. Oliver and Boyd, Edinburgh, 1935; 8th ed., 1966, reprinted by Hafner, New York, 1971. References are to the 5th ed., 1949, but section numbers are unchanged in successive editions.
- ST *Statistical Tables for Biological, Agricultural, and Medical Research* (with Frank Yates). Oliver and Boyd, Edinburgh, 1938; 6th ed., 1963.
- SI *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh, 1956; 3rd ed., Hafner, New York, 1973. Page references are to the 1st ed.
- CMS *Contributions to Mathematical Statistics*. Wiley, New York, 1950. Reprints of 42 statistical papers, selected with some emendations and prefatory notes by the author, and of Mahalanobis (1938). Index by John W. Tukey. Original pagination used, with a and b for prefatory notes.
- CP *Collected Papers of R. A. Fisher* 1 (1912–24), 2 (1925–31), 3 (1932–36), 4 (1937–47), 5 (1948–62). J. H. Bennett, ed., Univ. of Adelaide 1971, 1972, 1973, 1974, 1974. Complete papers and some other writings, with corrections and occasional comments, including the CMS emendations and prefatory notes. Original pagination shown. Vol. 1 contains a very complete bibliography of Fisher and the Biographical Memoir by Yates and Mather (1963). Vol. 5 contains autocitation, name, and subject indexes to all five volumes.
- 1912 On an absolute criterion for fitting frequency curves. *Messeng. Math.* **41** 155–160.
- 1913 Applications of vector analysis to geometry. *Messeng. Math.* **42** 161–178.
- 1914 Some hopes of a eugenicist. *Eugen. Rev.* **5** 309–315.
- 1915 Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10** 507–521.
- 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh* **52** 399–433.
- 1920 A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices Roy. Astronom. Soc.* **80** 758–770. Paper 2 in CMS.
- 1921a On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* **1** (4) 3–32. Paper 1 in CMS, Author’s Note only.
- 1921b Studies in crop variation. I. An examination of the yield of dressed grain from Broadbalk. *J. Agric. Sci.* **11** 107–135. Paper 3 in CMS.
- 1922a On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368. Paper 10 in CMS.

- 1922b On the interpretation of  $\chi^2$  from contingency tables, and the calculation of  $P$ . *J. Roy. Statist. Soc.* **85** 87–94. Paper 5 in CMS.
- 1922c The goodness of fit of regression formulae, and the distribution of regression coefficients. *J. Roy. Statist. Soc.* **85** 597–612. Paper 6 in CMS.
- 1922d The accuracy of the plating method of estimating the density of bacterial populations, with particular reference to the use of Thornton's agar medium with soil samples (with H. G. Thornton and W. A. MacKenzie). *Ann. Appl. Biol.* **9** 325–359. Paper 4 in CMS.
- 1923a Note on Dr. Burnside's recent paper on errors of observation. *Proc. Cambridge Philos. Soc.* **21** 655–658.
- 1923b Statistical tests of agreement between observation and hypothesis. *Economica* **3** 139–147. Paper 7 in CMS.
- 1923c Studies in crop variation. II. The manurial response of different potato varieties (with W. A. MacKenzie). *J. Agric. Sci.* **13** 311–320.
- 1924a The conditions under which  $\chi^2$  measures the discrepancy between observation and hypothesis. *J. Roy. Statist. Soc.* **87** 442–450. Paper 8 in CMS.
- 1924b On a distribution yielding the error functions of several well known statistics. *Proc. Internat. Cong. Math., Toronto* **2** 805–813. Paper 12 in CMS.
- 1924c The influence of rainfall on the yield of wheat at Rothamsted. *Philos. Trans. Roy. Soc. London Ser. B* **213** 89–142.
- 1925a Theory of statistical estimation. *Proc. Cambridge Philos. Soc.* **22** 700–725. Paper 11 in CMS.
- 1925b Note on the numerical evaluation of a Bessel function derivative (with P. R. Ansell). *Proc. London Math. Soc. Ser. 2* **24** 54–56.
- 1925c Sur la solution de l'équation intégrale de M. V. Romanovsky. *C. R. Acad. Sci., Paris* **181** 88–89.
- 1926a The arrangement of field experiments. *J. Ministry Agric. Great Brit.* **33** 503–513. Paper 17 in CMS.
- 1926c On the random sequence. *Quart. J. Roy. Meteorol. Soc.* **52** 250.
- 1928a The general sampling distribution of the multiple correlation coefficient. *Proc. Roy. Soc. Ser. A* **121** 654–673. Paper 14 in CMS.
- 1928b On a property connecting the  $\chi^2$  measure of discrepancy with the method of maximum likelihood. *Atti Congr. Internaz. Mat., Bologna* **6** 95–100. Paper 9 in CMS.
- 1928c Limiting forms of the frequency distribution of the largest or smallest member of a sample (with L. H. C. Tippett). *Proc. Cambridge Philos. Soc.* **24** 180–190. Paper 15 in CMS.
- 1928d Maximum- and minimum-correlation tables in comparative climatology (with T. N. Hoblyn). *Geogr. Ann.* **10** 267–281.
- 1929a Moments and product moments of sampling distributions. *Proc. London Math. Soc. Ser. 2* **30** 199–238. Paper 20 in CMS.

- 1929b Tests of significance in harmonic analysis. *Proc. Roy. Soc. Ser. A* **125** 54–59. Paper 16 in CMS.
- 1929c The sieve of Eratosthenes. *Math. Gaz.* **14** 564–566.
- 1929d Studies in crop variation. VI. Experiments on the response of the potato to potash and nitrogen (with T. Eden). *J. Agric. Sci.* **19** 201–213. Paper 18 in CMS.
- 1929f Statistics and biological research. *Nature* **124** 266–267.
- 1929g The evolution of dominance; reply to Professor Sewall Wright. *Amer. Natur.* **63** 553–556.
- 1929h Reviews of *The Balance of Births and Deaths*, vol. 1: *Western and Northern Europe* by R. R. Kuczynski and *The Shadow of the World's Future* by Sir G. H. Knibbs. *Nature* **123** 357–8.
- 1930a The moments of the distribution for normal samples of measures of departure from normality. *Proc. Roy. Soc. Ser. A* **130** 16–28. Paper 21 in CMS.
- 1930b Inverse probability. *Proc. Cambridge Philos. Soc.* **26** 528–535. Paper 22 in CMS.
- 1930c The distribution of gene ratios for rare mutations. *Proc. Roy. Soc. Edinburgh* **50** 205–220. Paper 19 in CMS.
- 1931a The derivation of the pattern formulae of two-way partitions from those of simpler patterns (with J. Wishart). *Proc. London Math. Soc. Ser. 2* **33** 195–208.
- 1931b The sampling error of estimated deviates, together with other illustrations of the properties and applications of the integrals and derivatives of the normal error function. *Math. Tables* **1** xxvi–xxxv. British Assoc. Adv. Sci. Paper 23 in CMS.
- 1932 Inverse probability and the use of likelihood. *Proc. Cambridge Philos. Soc.* **28** 257–261.
- 1933a The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. Roy. Soc. Ser. A* **139** 343–348.
- 1933b Review of *Tables for Statisticians and Biometricians* ed. by Karl Pearson. *Nature* **131** 893–4.
- 1934a Two new properties of mathematical likelihood. *Proc. Roy. Soc. Ser. A* **144** 285–307. Paper 24 in CMS.
- 1934c The  $6 \times 6$  Latin squares (with F. Yates). *Proc. Cambridge Philos. Soc.* **30** 492–507.
- 1934d Randomisation, and an old enigma of card play. *Math. Gaz.* **18** 294–297.
- 1935a The mathematical distributions used in the common tests of significance. *Econometrica* **3** 353–365. Paper 13 in CMS.
- 1935b The logic of inductive inference (with discussion). *J. Roy. Statist. Soc.* **98** 39–82. Paper 26 in CMS.
- 1935c The fiducial argument in statistical inference. *Ann. Eugen.* **6** 391–398. Paper 25 in CMS.

- 1935d The case of zero survivors in probit assays. *Ann. Appl. Biol.* **22** 164–165.
- 1935e Statistical tests. *Nature* **136** 474.
- 1935f Discussion of Neyman (1935). *J. Roy. Statist. Soc. Suppl.* **2** 154–157, 173.
- 1936a Uncertain inference. *Proc. Amer. Acad. Arts Sci.* **71** 245–258. Paper 27 in CMS.
- 1936b The “coefficient of racial likeness” and the future of craniometry. *J. Roy. Anthropol. Inst.* **66** 57–63.
- 1936c A test of the supposed precision of systematic arrangements (with S. Barbacki). *Ann. Eugen.* **7** 189–193. Paper 28 in CMS.
- 1936d Has Mendel’s work been rediscovered? *Ann. of Sci.* **1** 115–137.
- 1937a Moments and cumulants in the specification of distributions (with E. A. Cornish). *Rev. Inst. Internat. Statist.* **5** 307–322. Paper 30 in CMS.
- 1937b Professor Karl Pearson and the method of moments. *Ann. Eugen.* **7** 303–318. Paper 29 in CMS.
- 1937c On a point raised by M. S. Bartlett on fiducial probability. *Ann. Eugen.* **7** 370–375.
- 1938 Presidential address, Indian statistical conference. *Sankhyā* **4** 14–17.
- 1939a The comparison of samples with possibly unequal variances. *Ann. Eugen.* **9** 174–180. Paper 35 in CMS.
- 1939b The sampling distribution of some statistics obtained from nonlinear equations. *Ann. Eugen.* **9** 238–249. Paper 36 in CMS.
- 1939c A note on fiducial inference. *Ann. Math. Statist.* **10** 383–388.
- 1939d “Student”. *Ann. Eugen.* **9** 1–9.
- 1940a On the similarity of the distributions found for the test of significance in harmonic analysis, and in Stevens’s problem in geometrical probability. *Ann. Eugen.* **10** 14–17. Paper 37 in CMS.
- 1940b The precision of discriminant functions. *Ann. Eugen.* **10** 422–429. Paper 34 in CMS.
- 1941a The negative binomial distribution. *Ann. Eugen.* **11** 182–187. Paper 38 in CMS.
- 1941b The interpretation of experimental four-fold tables. *Science* **94** 210–211.
- 1941c The asymptotic approach to Behrens’s integral, with further tables for the *d* test of significance. *Ann. Eugen.* **11** 141–172.
- 1942a The likelihood solution of a problem in compounded probabilities. *Ann. Eugen.* **11** 306–307. Paper 42 in CMS.
- 1942b The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen.* **11** 341–353. Paper 39 in CMS.
- 1942c Some combinatorial theorems and enumerations connected with the numbers of diagonal types of a Latin square. *Ann. Eugen.* **11** 395–401. Paper 41 in CMS.
- 1943 Review of *The Advanced Theory of Statistics*, vol. 1 by Maurice G. Kendall. *Nature* **152** 431–432.

- 1945a A system of confounding for factors with more than two alternatives, giving completely orthogonal cubes and higher powers. *Ann. Eugen.* **12** 283–290. Paper 40 in CMS.
- 1945b The logical inversion of the notion of the random variable. *Sankhyā* **7** 129–132.
- 1945c A new test for  $2 \times 2$  tables. *Nature* **156** 388.
- 1948 Biometry. *Biometrics* **4** 217–219.
- 1950a A class of enumerations of importance in genetics. *Proc. Roy. Soc. Ser. B* **136** 509–520.
- 1950b *Creative aspects of natural law*. 4th Eddington Memorial Lecture. Cambridge Univ. Press.
- 1952a Sequential experimentation. *Biometrics* **8** 183–187.
- 1952b Statistical methods in genetics. *Heredity* **6** 1–12.
- 1953a Dispersion on a sphere. *Proc. Roy. Soc. Ser. A* **217** 295–305.
- 1953c The expansion of statistics. *J. Roy. Statist. Soc. Ser. A* **116** 1–6.
- 1954 Discussion of Fieller, Creasy and David (1954). *J. Roy. Statist. Soc. Ser. B* **16** 212–213.
- 1955 Statistical methods and scientific induction. *J. Roy. Statist. Soc. Ser. B* **17** 69–78.
- 1957 The underworld of probability. *Sankhyā* **18** 201–210.
- 1958a The nature of probability. *Centennial Rev.* **2** 261–274.
- 1958b Cigarettes, cancer, and statistics. *Centennial Rev.* **2** 151–166.
- 1958c Review of *The Autobiography of Charles Darwin*, ed. by Nora Barlow. *Nature* **182** 71.
- 1960 Scientific thought and the refinement of human reasoning. *J. Operations Res. Soc. Japan* **3** 1–10.
- 1961a Sampling the reference set. *Sankhyā Ser. A* **23** 3–8.
- 1961b The weighted mean of two normal samples with unknown variance ratio. *Sankhyā Ser. A* **23** 103–114.

## DISCUSSION

B. EFRON

*Stanford University*

This paper makes me happy to be a statistician. Savage, and John Pratt, give us a virtuoso display of high grade unobtrusive scholarship. Fisher comes across as a genius of the first rank, perhaps the most original mathematical scientist of the century. A difficult genius though, one in whom brilliance usually outdistances clarity. Savage deserves special credit for his deft pulling together of the various strings of Fisher's thought. This paper will make rereading Fisher easier for those of us raised in different statistical traditions.

My paper [1] is mainly devoted to understanding one of Fisher's inspired



guesses: that the maximum likelihood estimator is more efficient than other BAN estimators, and that the Fisher information measures variance of the MLE to a surprisingly high order of accuracy. Rao calls this first property of the MLE “second order efficiency.” The discussion following [1] contains some interesting comments on the reality or unreality of second order efficiency. All of this stems from two pages in Fisher’s 1925 paper (1925a). There are big issues at stake here. Fisher believes that the MLE is optimum as an information gathering statistic in finite samples, not just asymptotically. This belief is based on a deep geometric understanding of the estimation problem in parameterized subsets of high dimensional multinomial families. At this time, 50 years after the 1925 paper, nobody has successfully disputed Fisher’s claim, but the principle isn’t universally accepted, either. As Savage shows, you can find counterexamples to its strict interpretation (see the discussion following [1] for another one), but even these are surprisingly technical.

Fisher wrong can be just as interesting as Fisher right. It is hard to accept Fisher’s endorsement of the likelihood principle. Bad experience with estimating the mean vector of a multivariate normal distribution has destroyed confidence in the “invariance principle,” but this principle is itself just a weak consequence of the strict likelihood principle. Stein [2] has given a forceful counterexample. On the other hand, the likelihood principle is remarkably useful, or at least convenient, for dealing with problems in sequential analysis, and doesn’t seem to lead to disasters when applied to standard situations involving just a few parameters. Once again Fisher seems to have had a deep if not infallible insight into the nature of statistical inference.

Savage’s remarkable paper shows that “deep” beats “infallible” every time. We are badly in need of more statistical philosophers of the Fisher–Savage calibre.

#### REFERENCES

- [1] EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3** 1189–1242.
- [2] STEIN, C. (1962). A remark on the likelihood principle. *J. Roy. Statist. Soc. Ser. B* **125** 565–573.

CHURCHILL EISENHART

*National Bureau of Standards*

Jimmie’s Fisher Memorial Lecture “On Rereading R. A. Fisher” was the finest talk I ever heard on any aspect of statistics. His presentation held me spellbound throughout its entirety, and many friends to whom I have mentioned this tell me that they were equally entranced. Now that his wisdom and insight on Fisher has reached the printed page, I am sure that most of those who heard the original presentation and many others too will refer to it again and again in years to come.

BRUNO DE FINETTI

*University of Rome*

**1. Disconcerting inconsistencies.** It has been a great pleasure, for me, to receive and read this paper: it seemed almost like listening to the typical conversations of L.J.S. in which the subject under discussion became steadily broader and deeper owing to a series of little valuable discoveries (an example, a counter-example, a singular case, an analogy, a possible extension or a paradoxical one) that occurred to him, and were introduced into the discourse often following a short pause and a long “Ooooh...”. Of course, this pleasure was intimately mingled with the painful remembrance that the possibility of renewing such exciting meetings has been suddenly interrupted by his death.

Concerning R. A. Fisher in particular, I am indebted to Savage for all of the little understanding I have been able to attain about his viewpoint. Outside the difficulty (for me) of Fisher’s English, I was disconcerted by the alternation, in his writings, of assertions and remarks, some completely in agreement and some completely in disagreement with each one of the possible viewpoints about statistics, and in particular with my own viewpoint.

My uneasiness about understanding Fisher’s position has, perhaps, been definitely removed only by this posthumous message from Savage, particularly by the remarks about Fisher’s inconsistencies, explained (§§ 4.4–4.6, and elsewhere) by the conflict between his convictions about the necessity for conclusions in the form of “posterior probabilities” of a Bayes-like kind, and his preconception against admitting subjective prior probabilities, as well as his rejection (rightly) of “conventional” ones (like the uniform distribution in Bayes’ own approach for “repeated trials,” and similar choices, e.g. of “conjugate priors,” if done merely “for mathematical convenience”). It is but as an attempt—or a subterfuge—to escape such an inescapable dilemma, that he resorts to inventing an undefined name like “fiducial probability” or to suggesting the use of “likelihoods” as ersatz probabilities. This is, indeed, a wrong answer to the rightly perceived “absolute necessity to the advance of science” of attaining Bayes’ goal, whose abandonment he regards as “rather like an acknowledgment of bankruptcy” (§ 4.6).

Let me mention here a remark by L. J. S. (§ 4.6) concerning a more general impoverishment of statistical thinking which occurs when the Bayesian outlook is lost sight of: relevant circumstances are “sometimes overlooked by non-Bayesians, because they do not have easy ways to express (them).” Their language is too one-sided, hence poor and distorting.

**2. Preliminary personal impressions.** Fundamentally—it seems to me (if it is not too bold to judge a great man on a very limited knowledge)—Fisher had a strong intuitive insight into the many special problems he investigated, and mastery in dealing with them, but his unifying aims were limited to attempts at connecting fragmentary results rather than oriented toward the acquisition of a

general consistent view about probability and its use (as are the two opposed ones, of objectivists following the Neyman–Pearson school and of subjectivists like L.J. Savage).

One point has been especially surprising to me (as well as, it seems, to Savage himself: see § 4.3). Fisher was *not* a frequentist in the usual sense, but subscribed to a somewhat sophisticated and refined version of the so-called “classical” definition: the version where the set of  $N$  equally likely cases ( $M$  favorable, with  $M/N = p$ ) is supposed countless ( $N$  very large, practically infinite). The worst absurdity of the frequentist “definition” is so avoided: in a succession of drawings all sequences are possible (if  $p = \frac{1}{2}$ , with equal probability) and the frequency is no way *obliged* to tend to  $p$ , nor to any limit whatsoever. This view is not contradictory, provided one avoids really infinite sets where it would be meaningless to speak of a “ratio”  $p = M/N = \infty/\infty$  (maybe “denumerable”/“denumerable”?). Of course, for a subjectivist (in particular, for myself) this view is tenable only if  $p$  is previously adopted as the evaluation of the probability concerned, and the  $N$  cases are chosen so as to be considered subjectively equally likely.

The fundamental inconsistency (with many ramifications) is, however, the one mentioned previously, and it deserves to be discussed in more detail and depth.

**3. Critical diagnosis of the inconsistencies.** The fundamental dilemma, mentioned in Section 1 and related to conflicts between some of Fisher’s convictions and preconceptions, even if perhaps unique in its essence, has naturally many connections with features of scientific thinking on one side and with practical aspects and applications on the other. Let us add some summary remarks and some more comments.

The broadest picture of the issue may be presented by the contrast existing—in Fisher’s mind—between *inductive reasoning* (as a purely scientific method of thinking), and *inductive behavior* (as a practical guide for action, particularly for economic decisions). This is in full opposition to the views of both subjectivists and objectivists (of the Neyman–Pearson school), who agree in considering the two notions as but the theoretical and applied side of the same thing. To deny this appears as strange as maintaining that addition requires different operations if concerned with pure numbers or amounts of money.

Fisher’s attitude is, moreover, contradictory in itself: while denying indignantly that methods based on optimization of economic results are therefore optimal per se, in a scientific context also, he insists conversely that his own “scientific” methods ought to be applied to practical decisions in order to avoid undue costs and losses. How can one maintain that what is best for  $A$  is also best for  $B$ , but not conversely?

Moreover, the economic approach seems (if not rejected owing to aristocratic or puritanic taboos) the only device apt to distinguish neatly what is or is not contradictory in the logic of uncertainty (or probability theory). That is the

fundamental lesson supplied by Wald's notion of "admissibility," formally identical to that of "coherence" in the Bayesian-subjectivistic approach: probability theory and decision theory are but two versions (theoretical and practical) of the study of the same subject: uncertainty. It seems therefore an underassessment of theories of "inductive behavior" to deny (as Neyman does) that they embrace all of the theory of "inductive logic" too.

**4. Incommunicability and isolation.** Personal factors are relevant to understanding some of Fisher's preconceptions; so the "silhouette" of Fisher painted by L. J. S. (§ 2) is not only delightful but essential as aid to understanding some more or less strange fixed ideas and some contradictory attitudes in identical situations. For example: criticizing in a single case the application of randomization (that he so vigorously advocated) because (as always) the sample could be "bad-looking" (§ 4.4); being nervous about the term "point estimation," properly specifying something he was steadily employing (§ 3); hostility to inductive behavior, and to other views, which seems often to be "adopted simply to avoid agreeing with his opponents" (§ 2.3); aversion toward Wald's remark that the choice of what experiment to do is a "decision" (§ 2.6); emphasizing as a success that not all his so-called "fiducial distributions" coincide with "posterior distributions" (instead of recognizing this case as nonsensical) (§ 4.6); recourse to the "very puzzling" (L. J. S.) logic of "something unusual" since (as L. J. S. remarked) that happens in almost any experiment (e.g., when any individual case, one of which must happen, has probability  $2^{-100}$ ) (§ 4.7). Even more puzzling seems (at least to me) the answer (just an empty witticism) of Fisher to L. J. S. about such "logic of something unusual"; was he unable to explain reasonably his own attitude? Or did the reasons seem to him so obvious that the question was idle, perhaps was itself (for him) an empty witticism?

Some polemical attitudes—e.g., against "mathematicians" (§ 2.1)—were probably excessive, even if limitation to strictly deductive reasoning is often no less excessive. Fisher seems to be right, however, in maintaining (§ 2.5) that mathematical education should be, in part, also informative and illustrative, not restricted to the part that is prescribed to be transmitted rigorously by proofs.

Finally, and incidentally (since seemingly unrelated to Fisher), a remark by L. J. S. (obvious, but perhaps not usually emphasized as it should be): *sufficiency*, for a statistic, is a property which can hold on the hypothesis of a given model; it is always *relative* (in this sense), never absolute (§ 4.8).

**5. Utility of further discussions.** The main utility of Savage's paper on R. A. Fisher consists—it seems to me—in presenting a synthetical but penetrating view of the controversial questions arising from his work. And the conclusion seems to be—no matter whether one is more or less inclined to prize or deny Fisher's ideas and contributions—that many such questions are still open or, if not, deserve attention for clarification of fundamental or secondary questions about statistical thinking and statistical practice.

L. J. S. justifies his writing saying that: *Those who have already read in Fisher will agree that understanding him deeply is not easy, and they may be glad to hear the views of another* (§ 1.1). That was surely true for me in reading him, and I am sure it would be so for me and many others in reading other opinions too.

L. J. S. says: *The following pages consist largely of judgments and impressions: nothing can convert these subjective reactions completely into objective facts...* (§ 1.2). But the same holds for every other, so that the comparison may improve and probably approach such persuasions.

L. J. S. adds that, however: *...it has been an invaluable discipline for me to support each of them by specific citations with reasonable thoroughness* (§ 1.2). Unfortunately, this search was not completed before his death; some valuable work has been devoted by John W. Pratt; it would be highly welcomed if statisticians familiar with Fisher's work could recall, find, and communicate additional quotations.

Let us hope that this paper by L. J. S. about Fisher may give rise to clarifying discussions about the foundations and the applications of statistics: a field about which very much has been said and will be said maybe for ever and ever, but where we may at any rate attain some progress by concentrating efforts on such a wide but specific range of questions.

D. A. S. FRASER

*University of Toronto*

We owe Professor Jimmie Savage deep appreciation for his thorough and detailed review of R. A. Fisher's statistical research and publications. And we also owe Professor John W. Pratt substantial thanks for his painstaking job of editing the original manuscript into final form for the *Annals* and assembling the extensive references needed for the manuscript.

Certainly Professor Savage's statistical viewpoint, the Bayesian viewpoint, is very different from the R. A. Fisher viewpoint. On occasions we are reminded of this by parenthetical references in the review, and indeed Professor Savage makes reference to "a somewhat personal review." I feel that much additional credit goes to Professor Savage for the way the difference in viewpoint has not affected the assessment of the many contributions made by R. A. Fisher.

In Section 4.8 Professor Savage discusses the likelihood principle and notes that "...it does not seem to have been plainly stated by him [Fisher] until his last book [*Statistical Methods and Scientific Inference*]." I have not had the impression that Fisher's writings supported the likelihood principle and indeed the specific references made to his last book do not leave me with a feeling that Fisher in any considered way supported the principle. The likelihood principle was a prominent topic at statistics meetings around 1962, largely as a result of Allan Birnbaum's research interests. At that time it was reported that Fisher had been asked concerning the likelihood principle, that Fisher had enquired

what the likelihood principle was and then thoughtfully replied that he did not support it. Perhaps others can comment more authoritatively on this.

Professor Savage remarks that the likelihood “principle is still controversial, but [he believes] that it will come to be generally accepted.” Certainly from the Bayesian viewpoint there are no grounds for doubting the principle. And in general we have lacked examples with any force against the principle. Professor Mervyn Stone (1976) in a very recent paper discusses some Bayesian complications found with two examples. The first of these is an elaboration on Problem 11 in Lehmann (1959 24). This example has strong implications beyond the Bayesian viewpoint: it can be presented as a powerful example *against* the likelihood principle. Readers of the review of R. A. Fisher’s work will want to consider this example in Stone (1976).

Professor Savage refers to “many doubts and few unequivocal answers” in Fisher’s work. He also quotes Fisher: “I am still too often confronted by problems . . . to which I cannot confidently offer a solution, ever to be tempted to imply that finality has been reached . . . .” I think that this has several implications concerning Fisher’s research and deserves further comment. One important characteristic of Fisher was his ability to move into new areas of statistics, suggesting concepts and methods and deriving results. In a larger sense this avoided premature crystalization and conceptualization and left the theory open to modification and development. Often however he was taken at face value on some technical issue and the issue pursued meticulously. For example, the likelihood function as used in Fisher in contrast to the common and incorrect definition in most statistics texts. And the concept of sufficiency as used in Fisher in contrast to the extensive mathematical analyses of sufficiency most of which became superfluous with the general recognition around 1960 of the likelihood function statistic, a recognition that was in fact in Fisher’s earliest papers on sufficiency. In retrospect the openendedness of Fisher’s exploratory work deserves more positive than negative credits.

Professor Savage notes that “It would be more economical to list the few statistical topics in which he displayed no interest than those in which he did.” If we examine present day statistics, the fruitful, basic, and scientifically useful parts of present day statistics, and then assess which concepts and methods were proposed or developed by Fisher, we would obtain a clearer picture of the magnitude of his contributions. In some measure the concepts and methods mentioned in the review do this. But an overview shows that Fisher’s contributions constitute the central material of present day statistics.

#### REFERENCES

- LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.  
 STONE, MERVYN (1976). Strong inconsistency from uniform priors (with discussion). *J. Amer. Statist. Assoc.* **71** 114–125.

V. P. GODAMBE

*University of Waterloo*

This paper by the late Professor Savage will, I believe, provide valuable suggestions for rereading Fisher to many, as indeed it did to me. The paper does touch upon most of the important aspects of Fisher's work. The presentation throughout is admirably "balanced" and "objective." Of course it is not possible to aim at completeness in such a short paper. I would therefore restrict my comments to a couple of details.

At the end of Section 4.4, "randomisation" is discussed. Concerning this topic I find the following two statements by Fisher difficult to reconcile. In his earlier paper (1936 b 58, 59) Fisher says: "The simplest way of understanding quite rigorously, yet without mathematics, what the calculations of the test of significance amount to, is to consider what would happen if our two hundred actual measurements were written on cards, shuffled without regard to nationality, and divided at random into two new groups of a hundred each. . . . Actually the statistician does not carry out this very simple and very tedious process, but his conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method." On the other hand in *Statistical Methods and Scientific Inference* (SI 98) Fisher says, ". . . and whereas planned randomisation (1935-1953) is widely recognized as essential in the selection and allocation of experimental material, it has no useful part to play in the formation of opinion, and consequently in the tests of significance designed to aid the formation of opinion in the Natural Sciences." [The references in the quotation are to his *Design of Experiments*.]

Now the latter statement above is made by Fisher in relation to Bartlett's test and other tests which introduce a deliberate random element to arrive at conclusions. Such randomisation, I agree with Professor Savage (last but one paragraph of Section 4.4) "does constitute a sort of exclusion of pertinent evidence." But Savage further says: "However, there seems to me to be a very similar fault in all those applications of randomisation that Fisher so vigorously advocated." At least the type of randomisation mentioned in the first quotation of Fisher in the last paragraph seems to be free from this "fault." Briefly, it seems to assert that the significance level obtained on the basis of randomisation frequency would be nearly the same as that obtained on the assumption of normality, when the assumption (model) is *valid*. A paper demonstrating this more elaborately is due to Eden and Yates (1933). The randomisation here surely does *not* imply "exclusion of any pertinent evidence." A restatement and extension of the above logic underlying randomisation, in my opinion, is as follows. For testing some hypothesis we construct a test-statistic which is appropriate with respect to the underlying model or assumptions (e.g. independence, normality, etc.). Now when the hypothesis is true and the model is valid, the test-statistic has a specified *probability distribution*. A suitable randomisation can generate for the test-statistic



a *frequency distribution* which numerically (approximately) agrees with the probability distribution just mentioned. Now if the model is valid the randomisation is obviously superfluous for inference. But if the model is not valid—and only then—the frequency distribution generated by the randomisation can provide some “inference.” Usually the experimenter will postulate some model based on his background knowledge. (Naturally he would not like to forego all his knowledge.) At the same time he cannot be very sure about the validity of this model. Under these conditions it is necessary that the inference based on the *probability distribution* obtained from the model agrees with the inference based on the frequency distribution generated by the randomisation. Obviously there *cannot* be any loss of “pertinent evidence” due to such a use of randomisation. In the above considerations, the “model” can also include a prior distribution. Thus appropriate randomisation can enable an experimenter to use a model of uncertain validity. This uncertainty often more realistically concerns some *aspect* of (not the entire) model. Only that randomisation which corresponds to this aspect is justified above. For details, I would like to refer to two papers (1971, 1973) by Thompson and myself.

There is another purpose (not entirely unrelated with the one discussed above) for which Fisher recommended randomisation; and that is to establish “causation.” This is the situation in design of experiments. Here, even though I find Fisher’s recommendation scientifically convincing, at a deeper philosophical level, I agree with Professor Savage that some “pertinent evidence” is lost by randomisation. (In the above, I owe a reference to Professor Sprott.)

In Section 4.2 Professor Savage comments on Fisher’s conjecture that among all Fisher consistent estimates, the maximum likelihood estimate loses the least information (SI 157). Here Savage says, “The conjecture is false as stated, though there may be some way to reconstrue it that makes it true.” In this connection it is of interest to note that Fisher earlier (1935 b 45, 46; also SI 142, 143) put forward the concept of linear estimating equations to define Fisher consistency. With a generalisation removing linearity, if one attempts to obtain the optimum (in terms of variance and/or amount of information) *estimating equation* instead of *estimate*, simple mathematics yields the optimality/efficiency of the *m.l. equation* for finite samples. The relevant references in this connection are Godambe (1960) and Bhapkar (1972).

#### REFERENCES

- BHAPKAR, V. P. (1972). On a measure of efficiency of an estimating equation. *Sankhyā Ser. A* **34** 467–472.
- EDEN, T. and YATES, F. (1933). On the validity of Fisher’s *z*-test. *J. Agric. Sci.* **23** 6–16.
- GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208–1211.
- GODAMBE, V. P. and THOMPSON, M. E. (1971). Bayes, fiducial and frequency aspects of statistical inference in regression analysis in survey-sampling (with discussion). *J. Roy. Statist. Soc. Ser. B* **33** 361–390.

GODAMBE, V. P. and THOMPSON, M. E. (1973). Philosophy of survey-sampling practice. To appear in the proceedings of 1973 workshop on *Foundations of probability and statistics and statistical theories of science* held at London, Ontario.

I. J. GOOD

*Virginia Polytechnic Institute and State University*

0. I shall number the sections of my comments to agree with Savage's numbers. The reader will often need to refer back to see the relevance of my comments.

1. True history of science would depend on letters, lectures and other oral communication as well as on publications, as in recent work on the early history of quantum mechanics. The pretence that nothing exists if it is not published is unfortunate, for it discourages people from talking about their work. Perhaps one day the history of statistics in this century will be properly discussed, and Savage's essay is a substantial contribution to such a treatment.

Good mathematicians have always used scientific induction in their work; for example, Gauss "discovered" the prime number theorem and quadratic reciprocity and never proved the former. Pólya uses probability only qualitatively in his writings on plausible inference in pure mathematics. I think it can be regarded as quantitative, though usually imprecise, and can be combined with the principle of rationality (maximization of expected utility). It involves a modification of the axioms of probability (Good, 1950 49), and a dynamic interpretation of probability which is useful both for the philosophy of science (Good, 1968a, 1971a, 1973) and for computer chess (Good, 1967a, 1976). Dynamic partially ordered probability resolves the controversies between Fisherian and Bayesian statistics.

2.3. In 1951, I met Fisher in Cambridge, and he mentioned that he thought the best contribution he was then likely to make to genetics would be to teach it to Cambridge mathematical students, partly because he thought they were exceptionally capable. He went on to say that most of his clients were not in the same class. (See § 4.4 below.)

In a colloquium in Cambridge in November 1954, R. B. Braithwaite gave a talk on the minimax method and its implications for moral philosophy. In the discussion I said that the minimax method suffered from the disadvantage of all objectivistic methods, including those used in Fisherian statistics, namely that they necessarily ignore information so as to achieve apparent objectivity. Thereupon Fisher rose furiously, with a white face, and asked me to address my comments to the contents of the lecture. After the meeting he told Henry Daniels I was an "upstart" though previously he had told Donald Michie that he liked my 1950 book.

George Barnard told me a few years ago that Fisher was well aware of his own tendency to lose his temper, and that he regarded it as the bane of his life.

I felt better disposed to Fisher after that, although S. Vajda told me Fisher once referred to an influential school of American statisticians as “Americans with foreign sounding names.”

2.4. Allowing for the cases mentioned by Savage, and for others, Fisher was to various extents anticipated regarding likelihood, measure of amount of information, the use of generating functions in branching processes, and the analysis of variance (by Lexis and others: see RW § 20 and Heiss, 1968), yet his contribution was great. To be partly anticipated should detract little from solid contributions.

Was Fisher a Bayesian? See § 4.6 below.

2.5. Some faults in Fisher’s style were (i) ambiguous use of pronouns; (ii) the annoying but comically incorrect use of the expression “dependent from”; (iii) covering up. For example, in RW (7th ed. at least) he describes his “exact test” for 2 by 2 contingency tables, and omits to mention that it assumes the marginal totals convey no information about independence. Yet in his (1935b 48), he states this assumption. (When he says “If it be admitted,” he in effect is saying that it is a matter of judgment.)

Fisher’s style might have been better if he had circulated his manuscripts for suggestions as Savage did with his 1954 book.

2.6. Fisher was a Galton Professor in London, and an admirer of Galton. In GT 252 Fisher says that Galton “was aware that among these [titled families] the extinction of the title took place with surprising frequency,” but in that book he refers only to Galton’s *Hereditary Genius* and not to *Natural Inheritance* where Watson’s work appeared. So Fisher may not have been aware of Watson’s work. The use of generating functions in branching processes was discovered independently by Bienaymé (1845), Watson (1873), Woodward (1948), and myself (1948), so there is no reason to suppose that it was too difficult for Fisher!

The notion of interactions in multidimensional contingency tables seems to start with a personal communication from Fisher to Bartlett, regarding  $2 \times 2 \times 2$  tables, as acknowledged in Bartlett (1935). This notion, together perhaps with the semi-Bayesian approach to two-dimensional tables of Good (1956, rejected in 1953 in its original form), and Woolf (1955), was part of the prehistory of the loglinear model. The interactions were given a further philosophical boost when they were related to maximum entropy (Good, 1963).

The Wishart distribution should perhaps be called the Fisher–Wishart distribution, since the bivariate form was essentially due to Fisher (1915).

3. (i) Someone once said at a meeting of the RSS that the only sufficient statistic is the complete set of observations, because no model is certain.

(ii) When the number of parameters is not merely large but infinite, as in nonparametric estimation of a continuous density, ML is certainly inappropriate, but maximum “penalized likelihood” makes good sense, where the penalty depends on “roughness” (Good, 1971 b, Good and Gaskins, 1971, 1972). This can

be interpreted both in a non-Bayesian manner, and also as maximum posterior density in the shape of density functions.

4.1. Savage says “asymptotic variance is an appropriate index of the width of the posterior distribution for almost any practical loss function.” For an invariant generalization of Fisher’s information, allowing for loss functions, with an application, see Good (1968b, 1969, 1971a) and Good and Gaskins (1971, 1972).

4.4. In DOE § 27, the reference to bad judgment makes it clear that one of the purposes of randomization is to protect the experimenter against his own bad judgment (cf. my § 2.3). The objectivity and precision apparently obtained by randomization can be achieved by the device of a *Statistician’s Stooge* who alone knows the random design selected and who is shot if he reveals it (for example, Good, 1974 124).

4.6. First Fisher thought the fiducial argument could produce probabilities out of nothing. Then Jeffreys (1939) showed that in a few cases the results were the same as would be obtained by assuming (improper) priors. In SI 56, Fisher seems to say that the argument is essentially Bayesian. In Good (1965 81), I showed that the argument was incompatible with Bayesian methods, and in Good (1971a 139), I pinpointed the precise step where Fisher had brought in a hidden assumption. He had slipped up because he never used a notation for conditional probability.

4.7. By 1956 I think Fisher had moved close to a compromise with a Bayesian position. Such a compromise is possible because of a loose relationship between tail-area probabilities and Bayes factors (Good, 1950 94; 1967b; and Good and Crook, 1974). Re SI 71, the ratio is the Maximum Bayes Factor.

4.8. In c. 1941, I said to Barnard that, given two simple statistical hypotheses, the likelihood ratio obviously exhausted all the information because it is equal to the Bayes factor on the odds, and that we were using its logarithm, the weight of evidence, for sequential testing of hypotheses, the idea being due to Turing (who won the war). Barnard told me that he was curiously enough also using a similar method for quality control, but he did not agree that it exhausted all the information, because of the possibility of an incorrect model.

**Postscript.** I agree that Fisher should be read more, and scissors-and-paste books less.

#### REFERENCES

- BARTLETT, M. S. (1935). Contingency table interactions. *J. Roy. Statist. Soc. Suppl.* **2** 248–252.  
 GOOD, I. J. (1948). Private communication to M. S. Bartlett, June 5.  
 GOOD, I. J. (1950). *Probability and the Weighing of Evidence*. Griffin, London.  
 GOOD, I. J. (1956). On the estimation of small frequencies in contingency tables. *J. Roy. Statist. Soc. Ser. B* **18** 113–124.

- GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Ann. Math. Statist.* **34** 911–934.
- GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. MIT Press.
- GOOD, I. J. (1967 a). A five-year plan for automatic chess. *Machine Intelligence II* (E. Dale and D. Michie, eds.; Oliver and Boyd, Edinburgh), 89–118.
- GOOD, I. J. (1967 b). A Bayesian significance test for multinomial distributions. *J. Roy. Statist. Soc. Ser. B* **29** 399–431.
- GOOD, I. J. (1968 a). Corroboration, explanation, evolving probability, simplicity, and a sharpened razor. *British J. Philos. Sci.* **19** 123–143.
- GOOD, I. J. (1968 b). Utility of a distribution. *Nature* **219** 1392.
- GOOD, I. J. (1969). What is the use of a distribution? *Multivariate Analysis II* (P. R. Krishnaiah, ed.; Academic Press, New York), 183–203.
- GOOD, I. J. (1971 a). The probabilistic explication of information, evidence, surprise, causality, explanation, and utility. *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.; Holt, Rinehart and Winston, Toronto), 108–141.
- GOOD, I. J. (1971 b). Nonparametric roughness penalty for probability densities. *Nature Physical Science* **229** 29–30.
- GOOD, I. J. (1973). Explicativity, corroboration, and the relative odds of hypotheses. *Synthese* **30** 39–73.
- GOOD, I. J. (1974). Random thoughts about randomness. In *PSA 1972* (Boston Studies in the Philosophy of Science; D. Reidel: Dordrecht), 117–135.
- GOOD, I. J. (1976). Dynamic probability, computer chess, and the measurement of knowledge. In *Machine Representations of Knowledge* (E. W. Elcock and D. Michie, eds.), D. Reidel, Dordrecht.
- GOOD, I. J. and CROOK, J. F. (1974). The Bayes/non-Bayes compromise and the multinomial distribution. *J. Amer. Statist. Assoc.* **69** 711–720.
- GOOD, I. J. and GASKINS, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58** 255–277.
- GOOD, I. J. and GASKINS, R. A. (1972). Global nonparametric estimation of probability densities. *Virginia J. of Sci.* **23** 171–193.
- HEISS, K.-P. (1968). Lexis, Wilhelm. *International Encyclopedia of the Social Sciences* **9** 271–276.
- WOODWARD, P. M. (1948). A statistical theory of cascade multiplication. *Proc. Cambridge Philos. Soc.* **44** 404–412.
- WOOLF, BARNET (1955). On estimating the relation between blood groups and disease. *Ann. Human-Genetics* **19** 251–253.

O. KEMPTHORNE

*Iowa State University*

The Fisher Memorial Lecture of L. J. Savage was the finest statistical lecture I have heard in my whole life. There is no suggestion or requirement that Fisher lectures should be addressed to Fisher's own work, and this lecture was a surprise for many. It consisted of a review of the main thrusts of Fisher's life done with deep respect and reflected a tremendous effort to understand and place Fisher's contributions in the history of statistical ideas. It is tragic that Savage could not complete his oral presentation for publication. The statistical profession is deeply indebted to J. W. Pratt for a remarkable effort.

As Savage said, to read the whole of Fisher's work with some semblance of partial understanding is a huge task, one on which many have spent years. I surmise that the breadth and depth of that work will not be adequately appreciated for decades. I suggest that no individual of this century contributed fundamentally to so wide a variety of areas. I have always been impressed by Fisher's ability as a working mathematician. It seems that Fisher could tackle successfully and with deep originality almost any problem involving classical analysis, numerical analysis, probability calculus, or combinatorics. I regard him as a mathematician of the very highest order particularly in the dimension of creativity. Curiously enough, it seems that Fisher did nothing on strong asymptotic laws.

Fisher's ability in distribution theory was surely remarkable in the context of the times, and almost all of the distributional theory he worked out has become part of the intermediate knowledge of mathematical statistics. Savage communicated in his presentation the marvel of this effort. Fisher also became deeply fascinated by any combinatoric problem, and his work on experimental designs and in mathematical genetics in this direction boggles the mind. Fisher was highly original in multivariate analysis.

One aspect of Fisher's work which was touched on only briefly by Savage was Fisher's genetical effort. This would involve a lecture of similar dimensions to the present one. It is noteworthy, however, that Fisher was also the first to attack discrete stochastic processes by means of diffusion approximations via the Fokker-Planck differential equation (even though the first effort contained a foolish mistake).

The mysteries of Fisher's thought arise as soon as one turns away from the purely mathematical work which has stood the test of time except for a small number of minor errors.

It seems quite clear that Fisher never succeeded in communicating to anyone his idea of the nature of probability in spite of many efforts. I now find his 1956 book (SI) almost a total mystery. Fisher really did think that one could develop by logical reasoning a probability distribution for one's *knowledge* of a physical constant. It is clear, I think, that Fisher did not support any idea of belief probabilities of the type that Savage himself developed and presented so forcefully. The fiducial argument was to lead to some sort of logical probability which Fisher claimed could be verified, though he never gave an understandable idea of what he meant by verification, and specifically excluded the possibility of repeated measurements of the unknown constant.

Savage alluded, appropriately, to obscurity on what Fisher meant by "estimation." My guess is that he meant the replacement of the data by a scalar statistic  $T$  for the scalar parameter  $\theta$  which contained as much as possible of the (Fisherian) information on  $\theta$  in the data. But what one should do with an obtained  $T$  was not clear, though Fisher was obviously not averse at times to regarding  $T$  as an estimator of  $\theta$ . It is interesting, as Savage noted, that Fisher was the first to formulate the idea of exponential families in this connection.

Here, also, the fascinating question of ancillaries arises, and on this Fisher was most obscure. To some extent Fisher must be regarded as the initiator of estimation as a decision theory process, even though other writings suggest that he found this view offensive. I imagine that without fiducial inference Fisher would have found his views incoherent.

The work of Fisher abounds in curiosities. One which has struck me forcibly is the absence of any discussion of the relationship of Fisher's ideas on experimentation (DOE) to his general ideas on inference (SI). The latter book contains no discussion of ideas of randomization (except for the irrelevant topic of test randomization) which made DOE so interesting and compelling to investigators in noisy experimental sciences. Can the ideas on randomization and on parametric likelihood theory be fused into a coherent whole? I think not. In DOE Fisher convinces us of the desirability of randomization and unbiased (over randomizations) estimation of error, but then proceeds to the so-called analysis of covariance in which the unbiased estimation of error cannot be achieved.

I note that Savage applauded Fisher on factorial design, examining the relevant experimental factors simultaneously. But the prescriptions of Fisher work well only if interaction is small and lack of interaction is rare. With interaction, Fisher's analyses of variance lose much of their force. Fisher did not appreciate the role of nonadditivity and this came out in the 1935 Neyman discussion.

Savage discussed Fisher's ideas on statistical tests and was not able to obtain a coherent picture of Fisher's approach. It is important that the obscurities be recognized. Clearly Fisher regarded statistical tests as evidential in nature, but to say this is, perhaps, merely to replace one obscure idea by another no less obscure.

As regards likelihood, the origins in Fisher's own writing are quite obscure. In the early days it was a tool for point estimation but later it was elevated to a principle, again with deep mystery.

On fiducial inference, Fisher's early writings had a superficial transparency which convinced many of its correctness, and was thought to be the answer to the age-old problem of induction. But, *obviously*, Fisher was unable to convey his ideas to anyone, and, further, Fisher did not attach weight to the fact that fiducial calculations were possible only in a very limited set of conditions, quite inadequate for the broad purposes of science.

The upshot of all this can only be feelings of wonderment and puzzlement which Savage conveyed effectively, with respect, openness, and a highly sincere attempt to understand.

Will the statistical profession ever reach the status of nearly absolute acceptance or rejection of any of Fisher's ideas on inference? Or is the profession to retain forever a psychosis of not understanding Fisher and suspecting that it is stupid on that account?

The profession will be grateful for the indefinite future that L. J. Savage made such a fine effort to help.



STEPHEN M. STIGLER

*University of Wisconsin*

The youngest generation of mathematical statisticians may consider Savage's essay on Fisher a delightful introduction to the work of one of the great men of our profession, and older generations, including several veterans of the twentieth-century revolution in statistics, may find it a provocative reminder of battles not forgotten. But the essay poses a challenge to a historian of statistics.

In his discussion of Fisher's beliefs and attitudes about statistics, Savage implicitly raises a question: What was Fisher's place, his role, in the development of modern mathematical statistics? We may now be entering an era sufficiently distant from the time of Fisher's greatest works (and their attendant controversies) that a proper answer to this question will become feasible. Indeed, Savage hints at some possible answers, but a correct assessment of Fisher's work will not come easily. In part, this is because the development of statistics before Fisher's time is not really well understood, and in part it is because the proper placement of Fisher's works, particularly his researches in the design of experiments and in genetics, will not be accomplished without a deep and extensive study of nearly the entire range of a century of science, from Astronomy to Zoology, from Laplace to Weldon and Bateson.

While an answer to the question raised must be deferred to another time, I would like to make one point not mentioned by Savage, which, as far as I am aware, has not been noted by any other commentator on Fisher. The point is that it is to Fisher that we owe the introduction of parametric statistical inference (and thus nonparametric inference). While there are other interpretations under which this statement can be defended, I mean it literally—Fisher was principally responsible for the introduction of the word “parameter” into present statistical terminology!

The use of the term “parameter” in its present sense seems to date from Fisher's 1922 paper on the foundations of statistics. At the turn of the century, the term was in occasional usage in mathematics and physics, but in statistical literature published before 1922 its absence is nearly total. In mathematics, one specific meaning of *parameter* was (see the *Oxford English Dictionary*) the latus rectum of a conic section; that is, if  $y = (x - b)^2/p$  represents a parabola,  $p$  was called the *parameter* of the curve. Thus it is natural to expect, comparing this expression with the exponent in a normal density, that the term would appear in the statistical literature, if only in this restricted sense. Such appearances exist, but they are rare. I've found only a single instance of the word “parameter” in the works of Karl Pearson (once in 1894, in this restricted sense), and none in Student's works or even in Fisher's pre-1922 papers. Edgeworth used the term, but only rarely, and then to describe a scale factor (e.g.  $p^{\frac{1}{2}}$ ), usually assumed known. (When discussing scale parameters of normal distributions, he preferred the term “modulus”; he credited both terms to Bravais (1846 257).) I have

searched for “parameter” in texts on the theory of errors, a wide selection of *Biometrika* papers, the volumes of J. R. S. S. and J. A. S. A. for 1921, and the cited references in Fisher (1922a), and only found three instances of its use (by Edgeworth (1921), in a footnote, in his usual restricted sense; by Sheppard (1899), where the term is used for the scale parameter of a normal distribution, in a sense even more restrictive and closer to the strict geometrical meaning than Edgeworth’s; and by Yule, once in 1920, referring to a dummy variable of integration). And yet, in 1922a, Fisher used “parameter” a total of 57 times, in the general, modern sense!

While the introduction of a term may strike some as of minor or no importance to statistics, in this instance Fisher’s prolific use of “parameter” is symbolic of a subtle but important development of his predecessors’ concepts of families of probability distributions. Early workers in mathematical statistics had limited their attention to families of distributions where only location or scale parameters were unknown; these families were either too restricted to require, or the functional form of the density not sufficiently specified to permit, the full force of Fisher’s general theory of parametric inference. Later, Karl Pearson’s and Edgeworth’s general families of distributions were in fact *too* general to make concepts such as sufficiency and efficient estimation meaningful.

Fisher’s step, his narrowing of the focus of his attention to general parametric families of distributions where the dependence of the distribution upon a small number of parameters was smooth and regular, held the key to many of his greatest achievements. To cite just two examples: (1) The recognition of particular cases of the concepts of sufficiency can be found in the works of at least three men (Laplace, Simon Newcomb, Edgeworth) in the century before Fisher’s discovery, but the isolation and abstraction of the concept was impossible as long as statisticians followed Pearson and considered families of distributions that permitted different forms for different values of the defining constants; it required Fisher’s conceptual step, symbolized by the replacement of the “frequency constants” and “quaesita” of Pearson and Edgeworth by “parameters.” (2) It was Fisher’s understanding of parametric families of frequency curves that made possible both his formulation of the efficient method of maximum likelihood, and his correction of Karl Pearson’s errors on the degrees of freedom of  $\chi^2$ . (See Stigler (1973, 1975).)

Savage’s essay provides us with a much-needed road map of Fisher’s statistical contributions, but for a measure of Fisher’s influence on our field we need look no further than the latest issue of any statistical journal, and notice the ubiquitous “parameter.” Fisher’s concepts so permeate modern statistics, that we tend to overlook one of the most fundamental!

#### REFERENCES

- BRAVAIS, A. (1846). Analyse mathématique sur les probabilités des erreurs de situation d’un point. *Mémoires Acad. Sci. Paris, par Divers Savans* 9 255–332.  
 EDGEWORTH, F. Y. (1921). Molecular statistics. *J. Roy. Statist. Soc.* 84 71–89.

- SHEPPARD, W. F. (1899). On the application of the theory of error to cases of normal distribution and normal correlation. *Philos. Trans. Roy. Soc. London Ser. A* **192** 101-167.
- STIGLER, S. M. (1973). Laplace, Fisher, and the discovery of the concept of sufficiency. *Biometrika* **60** 439-445.
- STIGLER, S. M. (1975). The transition from point to distribution estimation. 40th Session of the I.S.I., Warsaw, Poland.

\* \* \* \* \*

It is possible to mention my appreciation for the effort of John Pratt in the preparation of this manuscript but I could never adequately describe my appreciation. To the many contributors to the final form of this manuscript I offer my sincere thanks.

I. RICHARD SAVAGE