Statistical Science
1992, Vol. 7, No. 3, 369–387

# R. A. Fisher and the Fiducial Argument

**S. L. Zabell**

*Abstract.* The fiducial argument arose from Fisher's desire to create an inferential alternative to inverse methods. Fisher discovered such an alternative in 1930, when he realized that pivotal quantities permit the derivation of probability statements concerning an unknown parameter independent of any assumption concerning its a priori distribution.

The original fiducial argument was virtually indistinguishable from the confidence approach of Neyman, although Fisher thought its application should be restricted in ways reflecting his view of inductive reasoning, thereby blending an inferential and a behaviorist viewpoint. After Fisher attempted to extend the fiducial argument to the multiparameter setting, this conflict surfaced, and he then abandoned the unconditional sampling approach of his earlier papers for the conditional approach of his later work.

Initially unable to justify his intuition about the passage from a probability assertion about a statistic (conditional on a parameter) to a probability assertion about a parameter (conditional on a statistic), Fisher thought in 1956 that he had finally discovered the way out of this enigma with his concept of *recognizable subset*. But the crucial argument for the relevance of this concept was founded on yet another intuition—one which, now clearly stated, was later demonstrated to be false by Buehler and Feddersen in 1963.

*Key words and phrases:* Fiducial inference, R. A. Fisher, Jerzy Neyman, Maurice Bartlett, Behrens-Fisher problem, recognizable subsets.

Most statistical concepts and theories can be described separately from their historical origins. This is not feasible, without unnecessary mystification, for the case of "fiducial probability." (Stone, 1983, p. 81)

## 1. INTRODUCTION

Fiducial inference stands as R. A. Fisher's one great failure. Unlike Fisher's many other original and important contributions to statistical methodology and theory, it has never gained widespread acceptance, despite the importance that Fisher himself attached to the idea. Instead, it was the subject of a long, bitter and acrimonious debate within the statistical community, and while Fisher's impassioned advocacy gave it viability during his own lifetime, it quickly exited the theoretical mainstream after his death.

Considerable confusion has always existed about the exact nature of the fiducial argument, and the entire subject has come to have an air of mystery about it. The root causes of such confusion stem from several factors. First and foremost, Fisher's own thoughts on fiducial inference underwent a substantial evolution over time, and both a failure on his part to clearly acknowledge this and a failure by others to recognize such changes have often led to confusion (when attempting to reconcile conflicting passages in Fisher's writings), or misinterpretation (when a later position is misread into an earlier). Second, fiducial inference never actually developed during Fisher's lifetime into a coherent and comprehensive theory, but always remained essentially a collection of examples, insights and goals, which Fisher added to and modified over time. Viewed in this limited way, the "theory" becomes at once much less ambitious and much more credible. Finally, the polemical nature of the debate on both sides rendered much of the resulting literature opaque: neither side was willing to concede inadequacies or limitations in its position, and this often makes any single paper difficult to understand when read in isolation.

This paper attempts to trace the roots and evolution of Fisher's fiducial argument by a careful examination of his own writings on the subject over a period of

*S. L. Zabell is Professor, Departments of Mathematics and Statistics, Northwestern University, 2006 Sheridan Road, Evanston, Illinois 60208.*

some thirty years. As will be seen, Fisher's initial insight and basic goals throughout are readily understood. But his attempts to extend the argument to the multiparameter setting and the criticism of his views by others led Fisher to reformulate the initial fiducial argument, and the approach taken in his later papers is very different from that to be found in his writings two decades earlier.

Although the last section of this paper briefly comments on the various efforts made after Fisher's death to clarify, systematize and defend the fiducial argument, our primary interest is what Fisher himself did (or did not) accomplish.

There are several "theses" advanced, stated below. These serve the useful purpose of summarizing the ensuing argument but necessarily omit a number of qualifications discussed later. Specifically, we will argue the following:

- Fisher's opposition to Bayesian methods arose (at least in part) from his break with Pearson; fiducial inference was intended as an "objective" alternative to "subjective," arbitrary Bayesian methods.
- Fisher's original fiducial argument was radically different from its later versions and was largely indistinguishable from the unconditional confidence interval approach later championed by Neyman.
- In response to Neyman's confidence interval formulation, Fisher drew attention to the multiplicity of conflicting parameter estimates arising from that approach, and in an attempt to deal with this difficulty he then explicitly imposed a further condition necessary for the application of the fiducial argument.
- As a result of his debate with Bartlett, Fisher became increasingly concerned with the conditional nature of inference, and this led to a dramatic shift in his conception of fiducial inference.
- Sensing that he was fighting a losing battle in the middle 1950s, Fisher made a supreme effort to spell out as clearly as he could the nature of the fiducial argument. In doing so, however, he revealed that the new intuitions he had about the fiducial argument were fundamentally incoherent.

## 2. FROM INVERSE TO FIDUCIAL PROBABILITY

Fisher began life a Bayesian. He tells us that, while at school, he learned the theory of inverse probability "as an integral part of the subject, and for some years saw no reason to question its validity" (Fisher, 1936, p. 248); he pled guilty to having, in his very first paper, "based my argument upon the principle of inverse probability" (Fisher, 1922, p. 326), and he thought it worth noting from an historical standpoint that "the ideas and nomenclature for which I am responsible were developed only after I had inured myself to the absolute rejection of the postulate of Inverse Probability" (CP 159A, p. 151).

Fisher saw fiducial inference as the jewel in the crown of the "ideas and nomenclature" for which he was responsible,[1] and in order to appreciate what he intended to achieve with it, we may perhaps best begin by considering what led him to so decisively reject the methods of inverse probability in the first place.

In 1915, Fisher published his first major paper in statistics, in which he derived the exact distribution of the sample correlation coefficient (Fisher, 1915). Although this paper was published in Karl Pearson's journal *Biometrika*, two years later a "cooperative study" by Pearson and several associates appeared criticizing Fisher's paper on several grounds (Soper et al., 1917). One of these, which particularly annoyed Fisher, was the (erroneous) charge that he had employed a Bayesian solution with an inappropriate prior for the correlation coefficient $\rho$.[2]

Relations between Fisher and Pearson rapidly worsened: by 1918 Pearson had rejected as referee Fisher's later famous paper on the correlation of relatives (Fisher, 1918), and the next year Fisher refused an offer from Pearson to join his laboratory, going to Rothamsted instead (Box, 1978, pp. 61, 82–83). Despite this, in 1920 Fisher again submitted a paper to Pearson for publication in *Biometrika*, but when this too was rejected Fisher vowed he would never do so again (Box, 1978, p. 83).[3]

Fisher's animosity toward Pearson is well known, but to gauge the true depth of his anger it is instructive to read the bitter Foreword to his book *Statistical Methods and Scientific Inference* (Fisher, 1956), written almost twenty years after Pearson's death.[4] It is at least arguable that in some cases the direction Fisher's statistical research now took – and the manner in which his papers were written – were motivated in part by a desire to attack Pearson. After moving to Rothamsted, Fisher proceeded (in a series of five papers published over the seven-year period 1922–1928) to attack Pearson's use of the chi-squared statistic to test homogeneity, on the (entirely correct) grounds that Pearson had systematically employed an incorrect number of degrees of freedom (Box, 1978; Feinberg, 1980). At the same time, Fisher began to publish his landmark papers on estimation. Although criticism of Pearson's work was not central to these, a key element of Fisher's new theory was the notion of efficient methods of estimation, and Fisher was quick to point out that Pearson's method of moments was frequently inefficient.

Pearson was also an exponent of Bayesian methods, and thus Fisher's rejection of inverse methods and his development of fiducial inference as an alternative to them was yet another assault on the Pearsonian edifice.[5] Less than a year after the rejection of his 1920 paper, Fisher fired the first salvo, asserting that the approach taken by Bayes "depended upon an arbitrary assumption, so that the whole method has been widely discredited" (Fisher, 1921, p. 4) and pointed an accusing finger at "inverse probability, which like an impenetrable jungle arrests progress towards precision of statistical concepts" (Fisher, 1922, p. 311).

Fisher could write with considerable conviction about the arbitrary nature of Bayesian analyses, for he felt that he had been one of its most recent victims. The writers of the *Cooperative Study*, Fisher charged, had altered

> my method by adopting what they consider to be a better *a priori* assumption as to $\rho$. This they enforce with such rigor that a sample which expresses the value 0.600 has its message so modified in transmission that it is finally reported as 0.462 at a distance of 0.002 only above that value which is assumed *a priori* to be most probable! (Fisher, 1921, p. 17)

The resulting value, Fisher thus noted, "depends almost wholly upon the preconceived opinions of the computer and scarcely at all upon the actual data supplied to him."

The close relationship between the *Cooperative Study* episode, and Fisher's subsequent and vehement rejection of inverse methods is evident in Fisher's original paper on fiducial inference (Fisher, 1930), appropriately called "Inverse Probability." Although the simplest examples of fiducial intervals would have been those for the mean and standard deviation, they were not employed by Fisher, who used instead the more complex example of the fiducial interval for the correlation coefficient, that is, precisely the setting in which Pearson had dared to criticize Fisher 13 years earlier for employing an inappropriate Bayesian solution. The slight had not been forgotten.

The exchange with Pearson impressed on Fisher the arbitrariness and dangers inherent in the use of priors lacking empirical support. By 1930, however, Fisher believed that he had discovered a way out of this difficulty, by employing what he termed the *fiducial argument.*

## 3. THE (INITIAL) FIDUCIAL ARGUMENT

### 3.1 The Birth of the Fiducial Argument

The fiducial argument was born during conversations between Fisher and his colleagues at Rothamsted.[6] A key role was played by the biologist E. J. Maskell, who worked there in the early 1920s. Maskell made the simple but important observation that when estimating the mean of a population, one could, in place of the usual two standard error limits, equally well employ the percentiles of the $t$-distribution to derive interval estimates corresponding to any desired level of significance.

Fisher briefly alluded to Maskell's role in Chapter 10 of the *Design of Experiments* (Fisher, 1935c). Referring to the classical example of Darwin's paired comparison of the heights of cross- and self-fertilized plants (introduced earlier in Chapter 3 of *Design of Experiments*, for which $n = 15$, $\bar{x} = 20.933$, $s = 37.744$, and $s/\sqrt{15} = 9.746$, Fisher wrote:

> An important application, due to Maskell, is to choose the values of $t$ appropriate to any chosen level of significance, and insert them in the equation. Thus $t$ has a 5 per cent. chance of lying outside the limits $\pm 2.154$. Multiplying this value by the estimated standard deviation, 9.746, we have 20.90 and may write
>
> $$\mu = 20.93 \pm 20.90$$
> $$= 0.03, \text{ or } 41.83$$
>
> as the corresponding limits for the value of $\mu$.[7]

But although Fisher thus knew the substance of the fiducial argument no later than 1926 (when Maskell left Rothamsted for the Caribbean), he did not refer to it in print for several years, perhaps because the initial observation, tied to the special case of the $t$-distribution, seemed too simple to warrant publication. But this changed by 1930, when Fisher discovered a way of generalizing the argument to cover a large class of univariate parameter estimates.

### 3.2 "Inverse Probability"

It is in many ways ironic that Fisher's first paper on fiducial inference, entitled "Inverse Probability" (Fisher, 1930), contains little that would be considered controversial today. In it Fisher introduced the probability integral transformation and observed that this transformation often provides a pivotal quantity which may be inverted to obtain interval estimates having any prespecified coverage frequency. That is, Fisher not only gave a clear and succinct statement of (what later came to be called) the confidence interval approach to parameter estimation, but (and this appears almost universally unappreciated) he also gave a general method for obtaining such estimates in the one-dimensional case.[8]

Fisher specifically observed that if a continuous statistic $T$ exists whose sampling distribution "is expressible solely in terms of a single parameter" $\theta$, then $T$ can often be inverted to obtain probability statements about $\theta$ which are true "irrespective of any assumption as to its *a priori* distribution":

> If $T$ is a statistic of continuous variation, and $P$

the probability that $T$ should be less than any specified value, we have then a relation of the form

$$P = F(T, \theta).$$

If now we give to $P$ any particular value such as 0.95, we have a relationship between the statistic $T$ and the parameter $\theta$, such that $T$ is the 95 per cent. value corresponding to a given $\theta$, and this relationship implies the perfectly objective fact that in 5 per cent. of samples $T$ will exceed the 95 per cent. value corresponding to the actual value of $\theta$ in the population from which it is drawn. To any value of $T$ there will moreover be usually a particular value of $\theta$ to which it bears this relationship; we may call this the "fiducial 5 per cent. value of $\theta$" corresponding to a given $T$. If, as usually if not always happens, $T$ increases with $\theta$ for all possible values, we may express the relationship by saying that the true value of $\theta$ will be less than the fiducial 5 per cent. value corresponding to the observed value of $T$ in exactly 5 trials in 100. By constructing a table of corresponding values, we may know as soon as $T$ is calculated what is the fiducial 5 per cent. value of $\theta$, and that the true value of $\theta$ will be less than this value in just 5 per cent. of trials. This then is a definite probability statement about the unknown parameter $\theta$ which is true irrespective of any assumption as to its *a priori* distribution. (Fisher, 1930, pp. 532–533)

That is, if $F(t, \theta) =: P_\theta[T \leq t]$, and if for each $p \in [0,1]$, the relation $F(t, \theta) = p$ implicitly defines functions $\theta_p(t)$ and $t_p(\theta)$ such that (i) $F(t_p(\theta), \theta) = p$ and (ii) $\theta_p(t) \leq \theta \Leftrightarrow t \leq t_p(\theta)$, then

$$P_\theta[\theta_p(T) \leq \theta] = p$$

whatever the value of $\theta$. Fisher termed $\theta_p(t)$ the "fiducial" $100(1 - p)$ percent value corresponding to $t$.[9]

This simple mathematical observation cannot, of course, be faulted, and all subsequent controversy about the fiducial argument has centered around either the *interpretation* of this result or the attempt to extend the argument to other contexts (discontinuous or multiparameter). Let us consider some of the issues raised either by Fisher or others.

### 3.3 The Interpretation of a Fiducial Probability

At this initial stage, Fisher's interpretation of the "fiducial" probability statement, as the quotation above makes clear, was closely tied to frequency considerations and coverage properties. Nor was the language occurring here an isolated instance, for it is closely paralleled by the language Fisher used in his next paper concerning the fiducial argument (Fisher, 1933).[10]

It might be argued that Fisher intended such references to frequency as simply stating a property (one among many) enjoyed by a fiducial interval, rather than being an essential element in its definition. Such an interpretation, however, is not supported by Fisher's language, for he went on to add (referring to the example of the correlation coefficient):

[I]f a value $r = 0.99$ were obtained from the sample, we should have a fiducial 5 per cent. $\rho$ equal to about 0.765. The value of $\rho$ can then only be less than 0.765 in the event that $r$ has exceeded its 95 per cent. point, an event which is known to occur just once in 20 trials. In this sense $\rho$ has a probability of just 1 in 20 of being less than 0.765. (Fisher, 1930, p. 534)

"In this sense"—this crucial phrase makes it clear that the references to sampling frequency that occur here and elsewhere were *central* to Fisher's conception of fiducial probability at this stage, not subsidiary to it.[11] The use of the adjective "fiducial," as Fisher himself repeatedly emphasized, was intended only to underscore the novel mode of derivation employed and was not meant to suggest that a new and fundamentally different type of probability was involved (as *was* the case with the distinction Fisher drew between probability and likelihood).[12]

### 3.4 The Fiducial Distribution

If the function $G(p) =: \theta_p(t)$ is strictly decreasing in $p$, then its inverse $G^{-1}(\theta)$ is certainly a distribution function in the mathematical sense that $H(\theta) =: 1 - G^{-1}(\theta)$ is a continuous increasing function with $H(-\infty) = 0$ and $H(+\infty) = 1$; Fisher termed it "the fiducial distribution" of the parameter $\theta$ corresponding to the value $t$ and noted that it has the density $-\partial F(t, \theta)/\partial \theta$.

Fisher regarded this result as supplying "definite information as to the probability of causes" and viewed the fiducial distribution as a probability distribution for $\theta$ in the ordinary sense. This is made clear at the end of the 1930 paper, when Fisher contrasted the fiducial and inverse approaches. At this stage, Fisher thought the fiducial argument was valid *even when* a prior distribution for $\theta$ was known. Because the resulting posterior and fiducial distributions ordinarily differ, Fisher stressed that they were really saying very different things: that although both were probability distributions, their "logical meaning" or "content" differed (Fisher, 1930, p. 534; 1933, pp. 82–83; 1936, p. 253), a position he later disavowed, for reasons that will be discussed below.[13]

Indeed Fisher's (1930) discussion (once again appealing to the correlation coefficient example he had used earlier) reveals just how unconditional a sampling interpretation he held at this juncture:

In concrete terms of frequency this would mean that if we repeatedly selected a population at random, and from each population selected a sample of four pairs of observations, and rejected all cases in which the correlation as estimated from the sample ($r$) was not exactly 0.99, then of the remaining cases 10 per cent. would have values of $\rho$ less than 0.765. Whereas apart from any sampling for $\rho$, we know that if we take a number of samples of 4, from the same or different populations, and for each calculate the fiducial 5 per cent. value for $\rho$, then in 5 per cent. of cases the true value of $\rho$ will be less than the value we have found. There is thus no contradiction between the two statements. (p. 535)

Little wonder that many statisticians during the 1930s regarded Fisher's theory of fiducial inference and Neyman's theory of confidence intervals as virtually synonymous![14] But despite the close similarities between the fiducial argument that Fisher presented in 1930 and Neyman's subsequent theory, there was—even at this early stage—an important difference in emphasis between the two. Every confidence interval is equivalent to a series of tests of significance, and it is clearly this second interpretation that Fisher had in mind. [Dempster (1964) expresses a similar philosophy, noting that "a particular 95% confidence region determined by observed data is simply the set of parameter values *not surprising* at the 0.05 level" (p. 58), and he suggests the term *indifference region* as more appropriate.][15] Fisher remained true to this interpretation, although his later analysis of just what constitutes a valid test of significance eventually led him to largely abandon the unconditional viewpoint adopted in these earlier papers.

To summarize thus far: for every fixed value of $\theta$, the statement $P_\theta[\theta_p(T) \leq \theta] = p$ has an unambiguous sampling interpretation for each $p$; for every fixed value of $t$, the function $H_t(\theta) =: 1 - F(t, \theta)$ is (in a purely mathematical sense) a distribution function for $\theta$. But Fisher did not regard the resulting fiducial distribution as a probability distribution for $\theta$ *in the sense that* it describes the frequency of $\theta$ in a population having fixed values of $T$;[16] the fiducial distribution of $\theta$ is only one in the sense that it is the "aggregate" of the probability statements $\{P_\theta[\theta_p(T) \leq \theta] = p\colon 0 \leq p \leq 1\}$ (each of which refers to the frequency of $T$ in a population having fixed values of $\theta$).[17] Fisher wrote in 1935:

> The [fiducial] *distribution* . . . is independent of all prior knowledge of the distribution of $\mu$, *and is true of the aggregate of all samples without selection*. It involves $\bar{x}$ and $s$ as parameters, but does not apply to any special selection of these quantities. (Fisher, 1935a, p. 392, emphasis added)

Thus, the fiducial *distribution* itself, and not just the individual probability statements $P_\theta[\theta_p(T) \leq \theta] = p$ which comprise it, must be interpreted in sampling terms. The point is that (for Fisher) every probability must be interpreted as a frequency in a population, and in this case the population is the one generated by repeated sampling: the "aggregate of all samples without selection." For $T = t$, one can compute the mathematical distribution function $H_t(\theta)$, but the probabilities in question do not refer to frequencies in a population where $t$ is fixed and $\theta$ variable.

Nevertheless, although the point is not discussed in his 1930 paper, Fisher did regard the fiducial distribution $H_t(\theta)$, given the observed sample value $T = t$, as a numerical measure of our rational degree of belief about different possible values of $\theta$ in the light of the sample, and on at least one occasion (although only in a letter), Fisher used the fiducial distribution for a fixed value of $t$ to compute distributional quantities such as the mean and median.[18] In order to understand this apparent conundrum, we need to pause to consider Fisher's concept of probability.

### 3.5 The Nature of Probability

Despite the straightforward nature of Fisher's 1930 paper, his language suggests the presence of more complex and potentially inconsistent views lurking beneath the surface. On the one hand, fiducial probabilities are defined in terms of objective, unconditional sampling frequencies; on the other, the fiducial argument is said to give rise to a "probability statement *about* the unknown parameter" (Fisher, 1930, p. 533; 1933, p. 82; 1935a, p. 391). The tension arises because for Fisher a probability is—by definition—a frequency in an infinite hypothetical population (Fisher, 1922), but it is also regarded by him as a "numerical measure of rational belief" (Fisher, 1930, p. 532; see also Fisher, 1935b, p. 40; Bennett, 1990, p. 121).

Fisher nowhere gives a systematic exposition of his pre-1940 views concerning probability, but its general outlines can be deduced from the scattered comments he makes throughout his papers.[19] For Fisher, probability has an objective value:[20] it is "a physical property of the material system concerned"[21] and is independent of our state of knowledge.[22] Numerically, it is a ratio of frequencies in an infinite hypothetical population,[23] that is, a mathematical limit of frequencies in finite populations.[24]

The process of statistical inference proceeds "by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample" (Fisher, 1922, p. 311). Such a population, being infinite, is necessarily imaginary, a mental construct: it is the "conceptual resultant of the conditions which we are studying" (1925, p. 700) and consists of the "totality of numbers produced by the same matrix of causal conditions" (1922, p. 313). Probability is defined in terms of hypothetical frequencies, not a limit of actual experimental frequencies, because we have no knowledge of the existence of such infinite experimental limits.[25] Nevertheless, experimental frequencies are an observational measure of probability, permitting their experimental verification (Fisher, 1934, p. 4).

Thus, for Fisher a probability is a frequency, an objective property of a specified population. But probability is also epistemic: it is the basis of inductive or uncertain inferences (Fisher, 1934, p. 6), plays a role in psychological judgment (1934, p. 287) and is a "numerical measure of rational belief" (Fisher, 1930, p. 532). This passage from a frequentist denotation to an epistemic connotation is the result of an unspecified process by means of which a class-frequency can be transferred from the class to an individual in that class (Fisher, 1935b):

> I mean by mathematical probability only that objective quality of the individual which corresponds to frequency in the population, of which the individual is spoken of as a typical member. (p. 78)

Thus, in the fiducial argument, given an observed value of $T$, say $t$, the probability statement concerning the parameter, $P[\theta_p(t) \leq \theta] = p$. is a numerical measure of our rational degree of belief concerning $\theta$, originating in the statement of objective frequency regarding the statistic $\theta_p(T)$, namely $P_\theta[\theta_p(T) \leq \theta] = P_\theta[T \leq t_p(\theta)]$, but then transferred after the observation of $T = t$ to the unknown and initially nonrandom parameter $\theta$. This can be found most clearly stated in a letter written much later to David Finney:

> The frequency ratio in the entire set, therefore, is the probability of the inequality being realized in any particular case, in exactly the same sense as the frequency in the entire set of future throws with a die gives the probability applicable to any particular throw in view. (Bennett, 1990, p. 98)

But as Fisher later came to realize in 1955, this passage from the frequency for a class to an epistemic probability for an individual indeed requires some justification. (Philosophers discuss this question under the rubric of the "problem of the single-case"; for example, Reichenbach and Salmon.)

In Fisher's writings, probability often seems to live a curious Jekyll and Hyde existence; for much of the time, probability leads a quiet and respectable life as an objective frequency (Dr. Jekyll), but it occasionally transforms before our very eyes into a rational degree of belief or even a psychological mental state (Mr. Hyde, of course). For most of us, the Jekyll and the Hyde peacefully coexist, but in Stevenson's tale, a crisis arises when the two begin to struggle for supremacy. Such a drama also occurred in the case of the fiducial argument.

## 4. NEYMAN AND CONFIDENCE INTERVALS

### 4.1 Neyman's 1934 *JRSS* Paper

Neyman left Poland at the beginning of 1934 in order to assume a permanent academic position at University College London. Shortly after his arrival in England, Neyman read a paper before the Royal Statistical Society (on 19 June 1934) dealing in part with the fiducial argument and reformulating Fisher's theory in terms of what Neyman called "confidence intervals" (Neyman, 1934).

After Neyman read his paper, one of the discussants who rose to comment on it was Fisher. The exchange between the two, taking place before relations between them broke down, is instructive. The tone was polite: in introducing his theory of confidence intervals, Neyman had described it as an alternative description and development of Fisher's theory of fiducial probability, permitting its extension to the several parameter case. Fisher, ironically one of the few to comment favorably on Neyman's paper, referred to Neyman's work as a "generalization" of the fiducial argument, but pointed to the problem of a possible lack of uniqueness in the resulting probability statements if sufficient or ancillary statistics were not employed and "the consequent danger of apparently contradictory inferences."[26]

Fisher began his discussion of fiducial inference (after briefly alluding to the question of terminology) by noting that his "own applications of fiducial probability had been severely and deliberately limited. He had hoped, indeed, that the ingenuity of later writers would find means of extending its application to cases about which he was still in doubt, but some limitations seemed to be essential" (p. 617).[27]

Fisher took it as a logical requirement of an inductive inference that it utilize all available information (here in the form of sufficient statistics or, lacking that, ancillaries), that probability statements not so based are necessarily deficient and that the multiplicity of possible interval estimates that could arise from Neyman's approach was, in effect, symptomatic of its failure to fully utilize the information in a sample. The rationale for the restriction to "exhaustive" statistics was thus *logical* rather than *mathematical*; that is, Fisher insisted on it not because it was necessary for the mathematical validity of the derivation, but because he viewed it as essential for the logical cogency of the resulting statement.

Confidence intervals, Fisher thought in contrast, make statements which, although mathematically valid, are of only limited inferential value. That they do indeed have *some* value was conceded by Fisher in a crucial footnote to his discussion:

> Naturally, no rigorously demonstrable statements, such as these are, can fail to be true. They can, however, only convey the truth to those who apprehend their exact meaning; in the case of fiducial statements based on inefficient estimates this meaning must include a specification of the process of estimation employed. But this process is known to omit, or suppress, part of the information supplied by the sample. The statements based on inefficient estimates are true, therefore, so long as they are understood not to be the whole truth. Statements based on sufficient estimates are free from this drawback, and may claim a unique validity. (pp. 617–618)[28]

In the remainder of his comments, Fisher made it clear that he did not view this problem as a minor one:

> Dr. Neyman claimed to have generalized the argument of fiducial probability, and he had every reason to be proud of the line of argument he had developed for its perfect clarity. The generalization was a wide and very handsome one, but it had been erected at considerable expense, and it was perhaps as well to count the cost. (p. 618)

Fisher then went on to list three specific reservations about Neyman's approach:

1. *The statistics employed were not restricted to those which were exhaustive.* Although Fisher had restricted the discussion in his 1930 paper to estimates arising from the method of maximum likelihood, the requirement there as stated is certainly cryptic, and in later years Fisher faulted his exposition for this reason.[29]

In a paper written shortly after, Fisher (1935a) remedied this omission by reviewing the logic of the fiducial argument in the case of a sample of size $n$ from a normal population with mean $\mu$. If $s_1$ denotes the sample standard deviation, $s_2$ the mean absolute deviation and

$$t_j = \frac{(\bar{x} - \mu)/\sqrt{n}}{s_j},$$

then, as Fisher noted, both $t_1$ and $t_2$ are pivotal quantities, and each can be employed to derive "probability statements" regarding the unknown parameter $\mu$, although in general the "probability distribution for $\mu$ obtained [from $t_2$] would, of course, differ from that obtained [from $t_1$]."

> There is, however, in the light of the theory of estimation, no difficulty in choosing between such inconsistent results, for it has been proved that, whereas $s_2$ uses only a portion of the information utilised by $s_1$, on the contrary, $s_1$ utilises the whole of the information used by $s_2$, or indeed by any alternative estimate. To use $s_2$, therefore, in place of $s_1$ would be logically equivalent to rejecting arbitrarily a portion of the observational data, and basing probability statements upon the remainder as though it had been the whole. (Fisher, 1935a, pp. 393–393)

2. *The extension to discontinuous variates was only possible by replacing an exact statement of fiducial probability by an inequality.* In particular, Fisher noted, "it raised the question whether exact statements of probability were really impossible, and if they were, whether the inequality arrived at was really the closest inequality to be derived by a valid argument from the data."

This clearly posed mathematical question interested Neyman, and his answer (largely negative) was published the next year (Neyman, 1935b). Fisher's own approach, characteristically clever, was unveiled in his 1935 Royal Statistical Society paper: in some cases a discontinuous variate can be transformed into a continuous variate amenable to the fiducial argument (Fisher, 1935a, pp. 51–53).[30] The problem of fiducial inference for discontinuous variates seems to have exercised a perennial fascination for Fisher; his obituary notice for "Student" gave in passing the simultaneous fiducial distribution for the percentiles of a continuous distribution by means of a discontinuous pivot (Fisher, 1939c, pp. 4–6), and he devoted a lengthy section to the problem of discontinuous variates in his book *Statistical Methods and Scientific Inference* (Fisher, 1956, pp. 63–70).[31]

3. *The extension to several unknown parameters.* Here, too, Fisher saw consistency as a major concern, contrasting the case of a single parameter, where "all the inferences might be summarized in a single probability distribution for that parameter, and that, for this reason, all were mutually consistent," with the multiparameter case, where it had not been shown that "any such equivalent frequency distribution could be established."

Neyman seems to have found this last reservation particularly puzzling,[32] but it clarifies Fisher's interest in the fiducial distribution as guaranteeing that the totality of inferential statements arising from the fiducial argument were mutually consistent.

Thus, Fisher's concerns at this stage were relatively straightforward. He insisted, on logical first principles, that the fiducial argument be limited to exhaustive statistics and saw the multiplicity of interval estimates that could arise from Neyman's approach as symptomatic of the failure of his theory to so limit itself.

### 4.2 The Break with Neyman

Although initially cordial, the relationship between Fisher and Neyman had never been warm, and in 1935, shortly after the above exchange, relations between the two broke down completely.[33] The occasion of the break was Fisher's discussion of Neyman's 1935 *Journal of the Royal Statistical Society* paper (read 28 March), which was sharply critical of Neyman, both in substance and tone.[34] Neyman's paper had itself been critical of some of Fisher's most important work, although the attack was indirect, and towards Fisher himself the tone of the paper is one of almost studied politeness. (This was not true, however, of Neyman's response.) The reasons for the pointedness of Fisher's attack can only be conjectured, but with it began a quarter-century long feud which dealt in part with fiducial probability, and we thus enter the second phase of Fisher's writings on the subject.[35]

But before going on to consider this phase, it is important to pause briefly and comment on the Fisher-Neyman dispute itself, because of a nearly universal misapprehension about its nature. Consider, for example, Neyman's description of the feud, summarized in his article "Silver Jubilee of My Dispute with Fisher":

> The first expressions of disapproval of my work were published by Fisher in 1935. During the intervening quarter of a century Sir Ronald honored my ideas with his incessant attention and a steady flow of printed matter published in many countries on several continents. All these writings, equally uncomplimentary to me and to those with whom I was working, refer to only five early papers, all published between 1933 and 1938. . . .
>
> Unfortunately, from the very start, [my dispute with Fisher] has been marred by Sir Ronald's unique style involving torrents of derogatory remarks. . . .
>
> Because of my admiration for the early work of Fisher, his first expressions of disapproval of my ideas were a somewhat shocking novelty and I did my best to reply and to explain. Later on, the novelty wore off and I found it necessary to reply only when Fisher's disapprovals of me included insults to deceased individuals for whom I felt respect. My last paper in reply to Fisher [appeared in 1956]. . . . Subsequent polemical writings of Fisher, including a book [Fisher, 1956], I left without reply. (Neyman, 1961, pp. 145–146, references omitted)

This undoubtedly reflected the way Neyman viewed the matter in 1961, but the picture it suggests is almost totally erroneous.

In reality, during the first two decades of the Fisher-Neyman dispute, far from "incessant attention," "a steady flow of printed matter" and "torrents of derogatory remarks," *Fisher almost never referred directly to Neyman in print.* For example, in the first ten years after their break (the period 1935–1944), Fisher referred to Neyman only twice in his papers (Fisher, 1935, 1941), and then only briefly.[36] Likewise, in the decade 1945–1954, one can only find two brief comments related to fiducial inference (Fisher, 1945, 1946); two brief asides in the *Collected Papers* (CP 204 and 205) unrelated to fiducial inference (one of which is innocuous) and a derogatory comment in *Contributions to Mathematical Statistics* (Fisher, 1950). In length, these five passages might comprise a total of two pages of text.

The Fisher-Neyman feud, of course, took place: the poisonous atmosphere in the University College Common room that their two groups shared is legendary. But initially it did not take place, for the most part, in print.[37] The one major exception (Neyman, 1941) was an attack on Fisher by Neyman and did not draw a response from Fisher. All this changed with the publication of Fisher's 1955 *Journal of the Royal Statistical Society* paper, and his 1956 book *Statistical Methods and Scientific Inference*, both of which repeatedly and sharply attacked Neyman in often highly uncomplimentary terms. But for the preceding twenty years of their feud, Fisher chose largely to *ignore* Neyman, and it is Fisher's 1955 paper and 1956 book, which Neyman identifies as the point when he, Neyman, withdrew from the fray, that in reality marks when Fisher's attack first began in earnest (for reasons that will be discussed below).

## 5. MULTIPARAMETER ESTIMATION

Neyman's claim to have gone beyond Fisher by developing methods for treating the case of several parameters must have seemed an obvious challenge. In a paper published soon after, Fisher presented an extension of the fiducial argument providing a solution to the problem of estimating the difference of two means, the so-called *Behrens-Fisher problem* (Fisher, 1935a; see, generally, Wallace, 1980).[38]

### 5.1 Fisher's 1935 Paper

Although Fisher emphasized in his 1935 paper (1935a) the necessity of using exhaustive estimates, he did not yet argue for the fiducial solution on the grounds of its conditional nature. Indeed, at one point, while comparing the Bayesian and fiducial approaches, Fisher actually stressed the *unconditional* nature of the fiducial argument:

> It is of some importance to distinguish [fiducial] probability statements about the value of $\mu$, from those that would be derived by the method of inverse probability. . . . The inverse probability distribution would specify the frequency with which $\mu$ would lie in any assigned range $d\mu$, by an absolute statement, true of the aggregate of cases in which the observed sample yielded the particu-

> lar statistics $\bar{x}$ and $s$. The [fiducial distribution] is independent of all prior knowledge of the distribution of $\mu$, *and is true of the aggregate of all samples without selection. It involves* $\bar{x}$ *and* s *as parameters, but does not apply to any special selection of these quantities.* (Fisher, 1935a, p. 392, emphasis added)

Thus, Fisher's conditional concerns did not arise from his dispute with Neyman but arose rather, as will be seen, because of his exchange with Bartlett.

Fisher's 1935 paper contains two important innovations that were to have a profound impact on the direction the fiducial debate later took. The first of these was the introduction of the *simultaneous fiducial distribution* (SFD); the second, the application of such distributions to *multiparameter estimation.* Maurice Bartlett, a young English statistician, soon raised important concerns about both of these innovations, and Bartlett's concerns, in one way or another, were to be at the heart of many of the later criticisms of fiducial inference. Let us consider each in turn.

### 5.2 The Simultaneous Fiducial Distribution

Fisher began by setting himself the problem of deriving a "unique simultaneous distribution" for the parameters of the normal distribution. The solution he proposed was ingenious. First illustrating how the fiducial argument could be employed, given a sample of size $n_1$ from a normal population with unknown $\mu$ and $\sigma$, to find the fiducial distribution of a single further observation (rather than, as before, unknown population parameters), Fisher showed how this approach could be generalized to obtain a fiducial distribution for the sample statistics $\bar{x}$ and $s$ arising from a second sample of size $n_2$, and then, by letting $n_2 \to \infty$, Fisher obtained a joint distribution for the population parameters $\mu$ and $\sigma$.

Where Fisher's 1930 paper had been cautious, careful and systematic, his 1935 paper was bold, clever but in many ways rash. For he now went on to conclude:

> In general, it appears that if statistics $T_1, T_2, T_3, \ldots$ contain jointly the whole of the information available respecting parameters $\theta_1, \theta_2, \theta_3, \ldots$, and if functions $t_1, t_2, t_3, \ldots$ of the T's and $\theta$'s can be found, the simultaneous distribution of which is independent of $\theta_1, \theta_2, \theta_3, \ldots$, then the fiducial distribution of $\theta_1, \theta_2, \theta_3, \ldots$ simultaneously may be found by substitution. (Fisher, 1935a, p. 395)

This sweeping claim illustrates the purely intuitive level at which Fisher was operating in this paper, and it was only towards the very end of his life that Fisher began to express doubts about this position.[39]

Fisher regarded the SFD as an ordinary probability distribution which could be manipulated in the usual ways, noting, for example, that the marginal distributions of the SFD for $(\mu, \sigma)$ were the previously known fiducial distributions for the two separate parameters. It was at this point that Fisher fell into a subtle trap; for in general the distribution of a function $f(\mu, \sigma)$ of the population parameters, induced by the SFD of $\mu$ and $\sigma$, will not satisfy the confidence property. The phenomenon already occurs, and is most easily understood, at the univariate level. If $X$ has a $N(\mu, 1)$ distribution, then the fiducial distribution for $\mu$ given $X = x$ is $N(x, 1)$, in the sense that if $P_\mu[\mu - X < c_\alpha] = \alpha$, then $P_\mu[\mu < X + c_\alpha] = \alpha$. If, however, the parameter of interest is $\mu^2$, the "fiducial distribution" for $\mu^2$ cannot be derived from that of $\mu$ in the usual way that the probability distribution for a random variate $U^2$ can be derived from that of $U$, if it is required that the limits arising from such a distribution satisfy the coverage property of Fisher's 1930 paper.[40]

This gap in Fisher's reasoning was later noted by Bartlett (1939), who pointed out that in the case of a normal sample the existence of the simultaneous distribution for $(\mu, \sigma)$ did not (for example) "imply that a fiducial inference could be made for . . . $\mu + \sigma$ by integration of the . . . fiducial distribution" (p. 133) and that, save in the very special case of the marginals of the SFD, "integration in any other problem is so far justified merely by analogy, and no statement as to its meaning in general has been given by Fisher" (p. 135). Bartlett's point here was completely correct, but his choice of example was exceedingly unfortunate, for it turns out that the *only* (!) univariate functions of $(\mu, \sigma)$ for which the confidence property is preserved when the SFD is integrated are precisely the linear functions $a\mu + b\sigma$ (see, e.g., Pedersen, 1978). Fisher pounced, and immediately pointed out the absence of any difficulty in the $\mu + \sigma$ example suggested by Bartlett (Fisher, 1941, pp. 143–146; see also 1956, pp. 125–127, 169).

Consistency questions such as these were basic to much of the fiducial debate in the 1950s, but at the time the ease with which Fisher answered Bartlett's specific question about the estimation of $\mu + \sigma$ may have seemed convincing enough to many. Bartlett's other objection to Fisher's multiparameter theory was not, however, so easily dealt with.

### 5.3 The Behrens-Fisher Problem

Fisher illustrated the uses of the simultaneous fiducial distribution with two examples, one of which was the notorious Behrens-Fisher problem. Few could have predicted then that it would generate a debate lasting several decades. Fisher's solution was almost immediately questioned by Bartlett (1936). Bartlett noted that, unlike the examples of the $t$-statistic, sample standard deviation and correlation coefficient, the interval estimates for $\mu_2 - \mu_1$ advocated by Fisher gave rise to tests with inappropriate levels of significance, in terms of frequencies involving repeated sampling from the same initial population. Although this must have been a rude surprise to Fisher, he quickly replied (Fisher, 1937)—the first in a series of exchanges with Bartlett over the next several years (Bartlett, 1937, 1939; Fisher, 1939a, 1939b, 1941; see also Bartlett, 1965).

Although in these exchanges Fisher professed to see no difficulty, he must in fact have been deeply troubled. It is revealing to read these papers as a group, for while Fisher kept returning to discuss the logic of the test, maintaining in public a confident air that all was well, the *grounds* on which this was asserted were constantly shifting.

Fisher rejected Bartlett's objection, initially (Fisher, 1937), on the not very convincing grounds that it introduced fixed values for the parameters into the argument, which Fisher argued was inconsistent with the assumed fiducial distribution. Fisher cannot have been comfortable with this response to Bartlett, because fixed values for the parameters had of course entered into his own original fiducial argument at one point.[41]

Two years later, when he returned to the question in response to another paper of Bartlett's (Bartlett, 1939; Fisher, 1939b), this defense was silently dropped, and Fisher defended his solution on the much more radical grounds that the very criterion being invoked by Bartlett was irrelevant:

> [T]he problem concerns what inferences are legitimate from a unique pair of samples, which supply the data, in the light of the suppositions we entertain about their origin; the legitimacy of such inferences cannot be affected by any supposition as to the origin of other samples which do not appear in the data. Such a population is really extraneous to the discussion. (p. 386)

This marked a major shift in Fisher's position.[42] Contrast, for example, Fisher's statement above with the language in his 1930 and 1935 papers cited earlier or, most strikingly, that in his 1933 paper:

> Probability statements of this type are logically entirely distinct from inverse probability statements, and remain true whatever the distribution *a priori* of $\sigma$ may be. To distinguish them from statements of inverse probability I have called them statements of fiducial probability. This distinction is necessary since the assumption of a given frequency distribution *a priori*, though in practice always precarious, might conceivably be true, in which case we should have two possible probability statements differing numerically, and expressible in a similar verbal form, though necessarily differing in their logical content. The probabilities differ in referring to different populations; that of the fiducial probability is the population of all possible random samples, that of the inverse probability is a group of samples selected to resemble that actually observed." (Fisher, 1933, p. 348)

Fisher's later writings tended to obscure this shift.[43] When Fisher republished a companion paper to the one above (Fisher, 1939a) in his 1950 collection *Contributions to Mathematical Statistics*, he singled out this point for comment in his introductory note:

> Pearson and Neyman have laid it down axiomatically that the level of significance of a test must be equated to the frequency of a wrong decision "in repeated samples from the same population." This idea was foreign to the development of tests of significance given by the author in 1925, for the experimenter's experience does not consist in repeated samples from the same population, although in simple cases the numerical values are often the same.... It was obvious from the first, and particularly emphasized by the present author, that Behrens' test rejects a smaller proportion of such repeated samples than the proportion specified by the level of significance, for the sufficient reason that the variance ratio of the populations sampled was unknown.

Such a statement is curiously inconsistent with Fisher's own earlier work. (See especially CP 48, pp. 503–505.) There is no hint in Fisher's 1934 contribution to the Neyman-Pearson theory of uniformly most powerful tests (Fisher, 1934) that he then considered their views to be "foreign to the idea of tests of significance," and when Fisher wrote to Neyman in 1932 commenting on the manuscript of the paper by Neyman and Pearson that later appeared in the *Philosophical Transactions* (Neyman and Pearson, 1933), his primary criticism was on a point of mathematical detail.[44] The assertion by Fisher that "the experimenter's experience does not consist in repeated samples from the same population" stands in contrast with the approach taken by him in his earliest papers on fiducial inference, where the argument is clearly cast in those terms. And far from it having been "obvious from the start" that "Behrens' test rejects a smaller proportion of such repeated samples," Fisher had explicitly conjectured in 1937 that this would *not* always be the case for samples of size greater than two.[45]

Fisher, of course, was certainly entitled to change his mind. But if only he had been willing to admit it![46]

In his papers of the 1930s, Fisher was just beginning to grapple with the problems of conditional inference, and his comments on these basic issues are at times brief, fragmentary, even tentative. It is symptomatic of the uncertainty he must have felt at this point that in 1941 he made the extraordinary concession that Jeffreys (1939, 1940), "whose logical standpoint is very different from my own, may be right in proposing that 'Student's' method involves logical reasoning of so novel a type that a new postulate should be introduced to make its deductive basis rigorous" (Fisher, 1941, p. 142).[47]

But when referring to Neyman, no such concession was possible. By 1945, Fisher's view had hardened, and he labeled the criterion that "the level of significance must be equal to the frequency with which the hypothesis is rejected in repeated sampling of any fixed population allowed by hypothesis" as an "intrusive axiom, which is foreign to the reasoning on which the tests of significance were in fact based" (Fisher, 1945, p. 507). Given the earlier frequency statements appearing in his first papers on fiducial inference, this was somewhat disingenuous.[48]

It is of course possible that Fisher's opposition to Neyman's "clarification" was based solely on an inability to accept that someone could improve on what he had already done.[49] But the evidence clearly suggests otherwise. Even in his discussion of Neyman's 1934 paper, Fisher had emphasized the necessity of utilizing all of the information in a sample; this was basic to Fisher's theory of statistical inference, pervasive in his earlier writings and implicit in his 1930 paper. Indeed, Fisher later claimed to have always insisted on it.[50] This was, moreover, precisely the time when Fisher was grappling with the difficulties of conditional inference, and in later exchanges Fisher would increasingly stress the importance in inference of conditioning on all relevant information. This is indeed a problem that the theory of confidence intervals has yet to resolve.

## 6. THE YEARS 1942-1955

That Fisher was uncomfortable with the theoretical underpinnings of his fiducial theory is suggested by the direction of his work during the next decade and a half: from 1942 to 1954 Fisher wrote almost nothing on fiducial inference, save a brief expository paper in *Sankhyā* (Fisher, 1945), a letter to *Nature* (Fisher, 1946), an expository paper in French (Fisher, 1948) and a discussion of a paper by Monica Creasy (Fisher, 1954).[51]

In his 1945 paper, Fisher illustrated the fiducial argument with the simple example of a sample of two drawn from a continuous distribution having a (presumably unique) median $\mu$.[52] If $X$ denotes the number of observations less than the median, then $X$ is a pivotal quantity with binomial distribution $B(2, 1/2)$; thus, for example, $P[\mu < \min(X_1, X_2)] = 1/4$. Fisher reasoned: "recognizing this property we may argue from two given observations, now regarded as fixed parameters that the probability is 1/4 that $\mu$ is less than both $x_1$ and $x_2$ . . . . The idea that probability statements about unknown parameters cannot be derived from *data* consisting of observations can only be upheld by those willing to reject this simple argument" (Fisher, 1945, p. 131).

This is candid enough, but it is really a complete admission of failure: it was precisely the cogency of this "simple argument" that Neyman (1941) and others had so vocally questioned.[53] Fisher could no longer appeal to the unconditional sampling justification of his earliest papers, but he was unable to supply an alternative given his new, conditional view of the matter. It was *intuitively* obvious to Fisher that the existence of a pivot warranted the transition from a probability assertion about statistics (conditional on the parameter) to a probability assertion about parameters (conditional on the statistic), but the significance of the passage quoted is that its language reveals that at this point Fisher was totally unable to supply further support or justification for that intuition.

This state of affairs lasted for a decade. Fisher's 1955 attack on the Neyman-Pearson approach to statistical inference in the *Journal of the Royal Statistical Society* (Fisher, 1955) touched only briefly on the specific question of fiducial inference, but, brief as his comments there are, they make it abundantly clear that he was no nearer to a satisfactory justification for the logical inversion central to the fiducial argument than he had been ten years earlier, when he wrote his expository piece for *Sankhyā*:

> A complementary doctrine of Neyman violating equally the principles of deductive logic is to accept a general symbolical statement such as
>
> $$Pr\{(\bar{x} - ts) < \mu < (\bar{x} + ts)\} = \alpha,$$
>
> as rigorously demonstrated, and yet, when numerical values are available for the statistics $\bar{x}$ and $s$, so that on substitution of these and use of the 5 per cent. value of $t$, the statement would read
>
> $$Pr\{92.99 < \mu < 93.01\} = 95 \text{ per cent.},$$
>
> to *deny* to this *numerical* statement any validity. This is to deny the syllogistic process of making a substitution in the major premise of terms which the minor premise establishes as equivalent. (p. 75)

A year later, however, in 1956, Fisher felt he had finally achieved a coherent rationale for the fiducial argument.

## 7. THE LAST BATTLE

Fisher's treatment of the fiducial argument in *Statistical Methods and Scientific Inference* (cited below as SMSI) (and nearly a dozen papers during the next several years) marks the third and final phase in his advocacy of fiducial inference. Perhaps realizing that he was now fighting a clearly downhill battle, Fisher made an obvious effort to present a clear statement of its logic.[54] Indeed, he went so far as to concede that "the applicability of the probability distribution [of the pivotal statistic] to the particular unknown value of [the parameter] . . . on the basis of the particular value of T given by his experiment, has been disputed, and ceertainly deserves to be examined" (SMSI, p. 57). This was a remarkable admission, since only a year earlier Fisher had excoriated Neyman for questioning precisely this applicability (Fisher, 1955, pp. 74–75)![55]

But in the interim Fisher's view of fiducial inference had radically altered. As he himself described it:

> It is essential to introduce the absence of knowledge *a priori* as a distinctive datum in order to demonstrate completely the applicability of the fiducial method of reasoning to the particular real and experimental cases for which it was developed. This point I failed to perceive when, in 1930, I first put forward the fiducial argument for calculating probabilities. For a time this led me to think that there was a difference in logical content between probability statements derived by different methods of reasoning. There are in reality no grounds for any such distinction. (SMSI, p. 59)

One might assume from Fisher's wording ("for a time this led me to think") that this shift in his thinking had occurred many years earlier. But that it had occurred only a short time earlier is evident from a passage in the *Design of Experiments* (1935c; see also 1951, p. 42). As late as the 6th revised edition of 1953 Fisher had continued to assert (emphasis added) the following:

> Statements of inverse probability have a different logical *content* from statements of fiducial probability, in spite of their similarity of form, *and* they require for their truth the postulation of knowledge beyond that obtained by direct observation. (Section 63)

But by 1956 Fisher no longer believed this, and thus in the next edition (7th, 1960) of *The Design of Experiments*, published a few years later, Fisher changed "content" to "basis" (as well as "and" to "for"). A basic and fundamental shift in Fisher's view of the nature of fiducial inference has thus been silently disguised by the subtle change of a single word. When Fisher wrote that inverse and fiducial statements differ in their *content*, he was referring primarily (at least in 1935, when this passage was first written) to the conditional aspect of the former and the unconditional aspect of the later. But when he says that they differ in their logical *basis*, he intends something quite different: both are, to use Dempster's phrase, "postdictive," but in one case based on prior knowledge (that is, a postulated prior distribution for the unknown parameter), in the other on the *absence* of prior knowledge of the parameter. But just exactly what does the verbal formulation "absence of prior knowledge" mean? Fisher had very early on rejected the Bayesian move that attempted to translate this into a uniform prior distribution for the parameter, for, as he noted (Fisher, 1922, p. 325), uniformity of prior is not invariant under parametric transformation. His insight in 1955–1956 was that the verbal, qualitative formulation "absence of prior knowledge" could be translated into an exact, quantitative postdictive distribution by invoking the fiducial argument—that the "absence of prior knowledge" was *precisely* the epistemological state which justified the invocation of the fiducial argument. This shift reflects in part his new view of the nature of probability and in part the device of recognizable subsets.

### 7.1 The Nature of Probability (continued)

Fisher's treatment of probability in SMSI reveals an apparent shift in his view of its nature. In his papers before World War II, Fisher had described prior distributions as referring to an objective process by which population parameters were generated. For example, writing in 1921, Fisher states that the problem of finding a posterior distribution "is indeterminate without knowing the statistical mechanism under which different values of [a parameter] come into existence" (Fisher, 1921, p. 24) and that "we can know nothing of the probability of hypotheses or hypothetical quantities" (p. 35). (In the 1950 introduction to this paper in *Contributions to Mathematical Statistics* (Fisher, 1950), Fisher brands the second assertion as "hasty and erroneous.")

In contrast, in the 1950s Fisher espoused a view of probability much closer to the personalist or subjectivistic one: "probability statements do not imply the existence of [the hypothetical] population in the real world. All that they assert is that the exact nature and degree of our uncertainty is just *as if* we knew [the sample] to have been one chosen at random from such a population" (Fisher, 1959, p. 22). None of the populations used to determine probability levels in tests of significance have "objective reality, all being products of the statistician's imagination" (Fisher, 1955, p. 71; cf. SMSI, p. 81). In the 1st and 2nd editions of SMSI, Fisher referred to "the role of subjective ignorance, as well as that of objective knowledge in a typical probability statement" (p. 33). [This embrace of the subjective was apparently too radical, however, for someone who had once tagged Jeffeys's system as "subjective and psychological" (CP 109, p. 3), and in the 3rd edition of SMSI, the passage was silently emended to read "the role both of well specified ignorance and of specific knowledge in a typical probability statement" (p. 35).]

Although Fisher remained publicly anti-Bayesian, after World War II he was in fact much closer to the "objective Bayesian" position than that of the frequentist Neyman.[56] In a little noted passage in SMSI, Fisher even cited without criticism Sir Harold Jeffreys's Bayesian derivation of the Behrens-Fisher interval, saying only that Jeffreys and others, recognizing "the rational cogency of the fiducial form of argument, and the difficulty of rendering it coherent with the customary forms of statement used in mathematical probability," had introduced "new axioms to bridge what was felt to be a gap," whereas "[t]he treatment in this book involves no new axiom" (p. 59). This was somewhat remarkable, inasmuch as Jeffreys's new axioms were modern reformulations of Bayes's postulate!

### 7.2 Recognizable Subsets

In Fisher's new view, an assertion of probability contained three elements: the specification of a reference set, the assertion that the outcome of interest was an element of this set and the assertion that no subset of the reference set was "recognizable" (SMSI, p. 60).[57] In the case of estimation, Fisher thought the absence of recognizable subsets a consequence of the requirements that the statistics employed be exhaustive and that there be absence of prior knowledge regarding the parameters (SMSI, p. 58). As an illustration, Fisher cited the case of the $t$-statistic:

> [T]he inequality
>
> $$\mu < \bar{x} - \frac{1}{\sqrt{N}} ts$$
>
> will be satisfied with just half the probability for which $t$ is tabulated, if $t$ is positive, and with the complement of this value if $t$ is negative. The reference set for which this probability statement holds is that of the values of $\mu$, $\bar{x}$, and $s$ corresponding to the same sample, for all samples of a given size of all normal populations. Since $\bar{x}$ and $s$ are jointly Sufficient for estimation, and knowledge of $\mu$ and $\sigma$ *a priori* is absent, there is no possibility of recognizing any sub-set of cases, within the general set, for which any different value of the probability should hold. (SMSI, p. 84; cf. Fisher, 1959, pp. 25–26)

This was clear enough, but Fisher's assertion, that the use of exhaustive estimates and the lack of knowledge a priori combined to insure the absence of recognizable subsets, was just that, an assertion. Seven years later Buehler and Feddersen (1963) somewhat unexpectedly showed that in precisely this case of the $t$-distribution recognizable subsets *did* exist, thus decisively refuting Fisher's final and clearest attempt at a justification.

A letter from Fisher to Barnard written in 1955 (14 October, Bennett, 1990, pp. 31–32) is revealing. Barnard, who had read a draft of this chapter in SMSI, queried Fisher about the justification for the fiducial distribution of $\mu$, and whether it was not based on the joint distribution of $\mu$ and $\sigma$. In reply, Fisher wrote that he did not think so, arguing as above:

> [I]f it is admitted that no subset can be recognized having a different probability, and to which the observed sample certainly belongs, (as can scarcely be disputed since $\bar{x}$ and s are jointly sufficient and it is postulated that no information *a priori* is available), the distribution of $\mu$ follows from that of $t$. (p. 32)

It is clear from Fisher's wording ("as can scarcely be disputed") that the basis for this assertion was an intuitive conviction on Fisher's part rather than a mathematical demonstration, and in a subsequent letter (on 17 October) a sceptical Barnard continued to query the point.

## 8. AFTERMATH

Although fiducial inference had its advocates in the years 1935–1955, a substantial majority of the statistical profession preferred the conceptual clarity of Neyman's confidence interval approach, and relatively few papers appeared on fiducial inference during this period. All this changed with the appearance of Fisher's book, which sparked renewed interest and controversy.

But that debate was largely possible only because of the ambiguities inherent in Fisher's theory (especially the method by which simultaneous fiducial distributions were to be constructed), his willingness in many instances to rely on intuition when asserting matters of mathematical fact and his preference for basing his treatment on "the semantics of the word 'probability'" (SMSI, p. 59), rather than axiomatics.[58]

Only in correspondence did Fisher express uncertainties never voiced publicly. As Fisher's friend George Barnard (1963) has noted, Fisher's "public utterances conveyed at times a magisterial air which was far from representing his true state of mind [regarding the fiducial argument in the case of several parameters]. In one letter he expresses himself as 'not clear in the head' about a given topic, while in another he referred ruefully to 'the asymptotic approach to intelligibility'" (p. 165). Indeed, Fisher once confessed to Savage, "I don't understand yet what fiducial probability does. We shall have to live with it a long time before we know what it's doing for us. But it should not be ignored just because we don't yet have a clear interpretation" (Savage, 1964, p. 926; see also Box, 1978, p. 458).

Once Fisher had gone from the scene, much of the heart went out of the fiducial debate, although important contributions continued to be made, most notably in the structural approach of Fraser and the belief function approach of Dempster. This literature was concerned not so much with fiducial inference, in the form Fisher conceived it, but with the attempt to achieve the goals for which it had been initially, if unsuccessfully, forged. As such it is beyond the scope of this paper. Three important papers which provide an entry into much of this literature are those of Wilkinson (1977), Pedersen (1978) and Wallace (1980).

The fiducial argument stands as Fisher's one great failure. Not only did he stubbornly insist on its cogency, clarity and correctness long after it became clear that he was unable to provide an understandable statement of it, let alone a coherent theory (Savage, 1976, p. 466, refers to Fisher's "dogged blindness about it all"), but he later engaged in a futile and unproductive battle with Neyman which had a largely destructive effect on the statistical profession. In SMSI, he was candid enough to confess the inadequacy of his earlier attempts to describe the fiducial argument and indiscreet enough to restate the argument with a clarity which permitted it to be decisively refuted.

Before his dispute with Neyman, Fisher had engaged in other statistical controversies, crossing swords with Arthur Eddington, Harold Jeffreys and Karl Pearson.[59] He had been fortunate in his previous choice of opponents: Eddington conceded Fisher's point, Jeffreys was cordial in rebuttal and Pearson labored under the disadvantage of being completely wrong.

But, in Neyman, Fisher was to face an opponent of an entirely different character.

## 9. CONCLUSION

The fiducial argument arose out of Fisher's desire to create an inferential alternative to inverse methods, avoiding the arbitrary postulates on which the classical Laplacean approach depended. Fisher felt he had discovered such an alternative in 1930, when he realized that the existence of pivotal quantities permitted the derivation of a probability distribution for an unknown parameter "irrespective of any assumption as to its *a priori* distribution" (p. 533).

The original fiducial argument, for a single parameter, was virtually indistinguishable from the confidence approach of Neyman, although Fisher thought its application should be restricted in ways that reflected his view of the logical basis of inductive reasoning. This effectively blended both an inferential and a behaviorist viewpoint. When Fisher subsequently attempted to extend the fiducial argument to the multiparameter setting in his treatment of the Behrens-Fisher problem, this conflict surfaced, and, forced to decide between the two, Fisher opted for the inferential, rather than the behaviorist route, thus (silently) abandoning the unconditional sampling approach of his earlier papers for the conditional approach of his later work.

Initially unable to justify his intuition about the passage from a probability assertion about a statistic (conditional on a parameter) to a probability assertion about a parameter (conditional on a statistic), Fisher thought in 1956 that he had finally discovered the way out of this enigma with his concept of *recognizable subset*. But despite the authoritative way in which Fisher asserted his new position in his last book, *Statistical Methods and Scientific Inference*, the crucial argument for the relevance of this concept was founded on yet another intuition – one which, now clearly stated, was later demonstrated to be false by Buehler and Feddersen in 1963.

Fiducial inference in its final phase was in essence an attempt to construct a theory of conditional confidence intervals (although Fisher would never have put it that way) and thereby "make the Bayesian omelette without breaking the Bayesian eggs." Fisher's failure, viewed in this light, was hardly surprising: no satisfactory theory of this type yet exists. But Fisher's attempt to steer a path between the Scylla of unconditional, behaviorist methods which disavow any attempt at "inference" and the Charybdis of subjectivism in science was founded on important concerns, and his personal failure to arrive at a satisfactory solution to the problem means only that the problem remains unsolved, not that it does not exist.

## ACKNOWLEDGMENT

At various stages in the drafting of this paper, I have received valuable and generous assistance and comments from many people interested in Fisher and his work. These include George Barnard, Maurice Bartlett, Arthur Dempster, Anthony Edwards, Erich Lehmann, Paul Meier, Teddy Seidenfeld and David Wallace. I am very grateful to them all.

## REFERENCES

Papers of Fisher referred to only on a single occasion are cited by their number in Fisher's *Collected Papers* (CP; Bennett, 1971–1974) and are not included below.

Barnard, G. (1963). Fisher's contributions to mathematical statistics. *J. Roy. Statist. Soc. Ser. A* **126** 162–166.

Barnard, G. (1990). Fisher: A retrospective. *Chance* **3** 22–28.

Bartlett, M. S. (1936). The information available in small samples. *Proceedings of the Cambridge Philosophical Society* **32** 560–566.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Statist. Soc. Ser. A* **160** 268–282.

Bartlett, M. S. (1939). Complete simultaneous fiducial distributions. *Ann. Math. Statist.* **10** 129–138.

Bartlett, M. S. (1965). R. A. Fisher and the first fifty years of statistical methodology. *J. Amer. Statist. Assoc.* **60** 395–409.

Bennett, J. H., ed. (1971–1974). *Collected Papers of R. A. Fisher.* Univ. Adelaide.

Bennett, J. H., ed. (1990). *Statistical Inference and Analysis:*

*Selected Correspondence of R. A. Fisher.* Clarendon Press, Oxford.

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

BUEHLER, R. J. and FEDDERSEN, A. P. (1963). Note on a conditional property of Student's t. *Ann. Math. Statist.* **34** 1098–1100.

CMS. See FISHER, 1950.

CREASEY, M. A. (1954). Limits for the ratio of means. *J. Roy. Statist. Soc. Ser. B* **16** 186–194.

DEMPSTER, A. P. (1964). On the difficulties inherent in Fisher's fiducial argument. *J. Amer. Statist. Assoc.* **59** 56–66.

EDWARDS, A. W. F. (1974). The history of likelihood. *Internat. Statist. Rev.* **42** 9–15.

FIENBERG, S. E. (1980). Fisher's contributions to the analysis of categorical data. *R. A. Fisher: An Appreciation. Lecture Notes in Statist.* **1** 75–84. Springer, New York.

FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10** 507–521. (CP 4.]

FISHER, R. A. (1918). On the correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* **52** 399–433. [CP 9.]

FISHER, R. A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* **1** 3–32. [CP 14.]

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368. [CP 18.]

FISHER, R. A. (1925). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh. [Many later editions.]

FISHER, R. A. (1930). Inverse probability. *Proceedings of the Cambridge Philosophical Society* **26** 528–535. [CP 84.]

FISHER, R. A. (1933). The concepts of inverse probability and fiducial probability referring to unknown parameters. *Proc. Roy. Soc. London Ser. A* **139** 343–348. [CP 102.]

FISHER, R. A. (1934). Two new properties of mathematical likelihood. *Proc. Roy. Soc. London Ser. A* **144** 285–307. [CP 108.]

FISHER, R. A. (1935a). The fiducial argument in statistical inference. *Annals of Eugenics* **6** 391–398. [CP 125.]

FISHER, R. A. (1935b). The logic of inductive inference (with discussion). *J. Roy. Statist. Soc.* **98** 39–82.

FISHER, R. A. (1935c). *The Design of Experiments.* Oliver and Boyd, Edinburgh. [Many later editions.]

FISHER, R. A. (1936). Uncertain inference. *Proceedings of the American Academy of Arts and Science* **71** 245–258. [CP 137.]

FISHER, R. A. (1937). On a point raised by M. S. Bartlett on fiducial probability. *Annals of Eugenics* **7** 370–375. [CP 151.]

FISHER, R. A. (1939a). The comparison of samples with possibly unequal variance. *Annals of Eugenics* **9** 174–180. [CP 162.]

FISHER, R. A. (1939b). A note on fiducial inference. *Ann. Math. Statist.* **10** 383–388. [CP 164.]

FISHER, R. A. (1939c). "Student." *Annals of Eugenics* **9** 1–9. [CP 165.]

FISHER, R. A. (1941). The asymptotic approach to Behrens's integral, with further tables for the *d* test of significance. *Annals of Eugenics* **11** 141–172. [CP 181.]

FISHER, R. A. (1945). The logical inversion of the notion of the random variable. *Sankhyā* **7**, 129–132. [CP 203.]

FISHER, R. A. (1946). Testing the difference between two means of observations of unequal precision. *Nature* **158** 713. [CP 207.]

FISHER, R. A. (1948). Conclusions fiduciaires. *Ann. Inst. H. Poincaré* **10** 191–213. [CP 222.]

FISHER, R. A. (1950). *Contributions to Mathematical Statistics* [CMS]. Wiley, New York.

FISHER, R. A. (1951). Statistics. In *Scientific Thought in the Twentieth Century* (A. E. Heath, ed.) 31–55. Watts, London. [CP 242.]

FISHER, R. A. (1954). Contribution to a discussion of a paper on interval estimation by M. A. Creasy. *J. Roy. Statist. Soc. Ser. B* **16** 212–213.

FISHER, R. A. (1955). Statistical methods and scientific induction. *J. Roy. Statist. Soc. Ser. B* **17** 69–78. [CP 261.]

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference* [SMSI]. Hafner Press, New York. [2nd ed., 1959; 3rd ed., 1973; Page references in the text are to the 3rd ed.]

FISHER, R. A. (1958). the nature of probability. *Centennial Review* **2** 261–274. [CP 272.]

FISHER, R. A. (1959). Mathematical probability in the natural sciences. *Technometrics* **1** 21–29. [CP 273.]

FISHER, R. A. (1960). Scientific thought and the refinement of human reasoning. *J. Oper. Res. Soc. Japan* **3** 1–10. [CP 282.]

GOOD, I. J. (1971). Reply to Professor Barnard. In *Foundations of Statistical Inference* (V. P. Godambe and D. A. Sprott, eds.) 138–140. Holt, Rinehart, and Winston, Toronto.

HACKING, I. (1990). *The Taming of Chance.* Cambridge Univ. Press.

JEFFREYS, H. (1932). On the theory of errors and least squares. *Proc. Roy Soc. London Ser. A* **138** 38–45.

JEFFREYS, H. (1939). *Theory of Probability.* Clarendon Press, Oxford. [2nd ed., 1948; 3rd ed., 1961.]

JEFFREYS, H. (1940). Note on the Behrens-Fisher formula. *Annals of Eugenics.* **6** 391–398.

KENDALL, M. G. (1963). Ronald Aylmer Fisher, 1890–1962. *Biometrika* **50** 1–15.

LANE, D. (1980). Fisher, Jeffreys, and the nature of probability. *R. A. Fisher: An Appreciation. Lecture Notes in Statist.* **1** 148–160. Springer, New York.

NEYMAN, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc. Ser. A* **97** 558–625.

NEYMAN, J. (1935a). Statistical problems in agricultural experimentation (with K. Iwaszkiewicz and St. Kolodziejczyk). *J. Roy. Statist. Soc. B Suppl.* **2** 107–180.

NEYMAN, J. (1935b). On the problem of confidence intervals. *Ann. Math. Statist.* **6** 111–116.

NEYMAN, J. (1941). Fiducial argument and the theory of confidence intervals. *Biometrika* **32** 128–150.

NEYMAN, J. (1961). Silver jubilee of my dispute with Fisher. *J. Oper. Res. Soc. Japan* **3** 145–154.

NEYMAN, J. and PEARSON, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Phil. Trans. Roy Soc. Ser. A* **231** 289–337.

PEARSON, E. S. (1968). Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments. *Biometrika* **55** 445–457.

PEARSON, E. S. (1990). *'Student': A Statistical Biography of William Sealy Gosset* (R. L. Plackett and G. A. Barnard, eds.). Clarendon Press, Oxford.

PEARSON, K. (1892). *The Grammar of Science.* Walter Scott, London. [2nd ed., 1900; 3rd ed., 1911.]

PEARSON, K. (1920). The fundamental problem of practical statistics. *Biometrika* **13** 1–16.

PEDERSEN, J. G. (1978). Fiducial inference. *Internat. Statist. Rev.* **46** 147–170.

REID, C. (1982). *Neyman—From Life.* Springer, New York.

SAVAGE, L. J. (1964). Discussion. *Bull. Inst. Internat. Statist.* **40** 925–927.

SAVAGE, L. J. (1976). On re-reading R. A. Fisher (with discussion). *Ann. Statist.* **4** 441–500.

SMSI. See FISHER, 1956.

SOPER, H. E., YOUNG, A. W., CAVE, B. H., LEE, A. and PEARSON, K. (1917). A cooperative study. On the distribution of the correlation coefficient in small samples. Appendix II to the Papers of 'Student' and R. A. Fisher. *Biometrika* **11** 328-413.

STONE, M. (1983). Fiducial probability. *Encyclopedia of Statistical Sciences* **3** 81-86. Wiley, New York.

WALLACE, D. (1980). The Behrens-Fisher and Fieller-Creasey problems. *R. A. Fisher: An Appreciation. Lecture Notes in Statist.* **1** 119-147. Springer, New York.

WILKINSON, G. N. (1977). On resolving the controversy in statistical inference (with discussion). *J. Roy. Statist. Soc. Ser. B* **39** 119-171.

WILKS, S. S. (1938). Fiducial distributions in fiducial inference. *Ann. Math. Statist.* **9** 272-280.

YATES, F. (1939). An apparent inconsistency arising from tests of significance based on fiducial distributions of unknown parameters. *Proceedings of the Cambridge Philosophical Society* **35** 579-591.

## ENDNOTES

1. See his statement in Fisher (1956, p. 77), in CP 290, Fisher contrasts the results of the fiducial argument with "those weaker levels of uncertainty represented by Mathematical Likelihood, or only by tests of significance."
2. Soper et al. (1917). After erroneously stating that "[Fisher] holds that *a priori* all values of $\rho$ are equally likely to occur" (p. 353), the authors discussed the consequences of assuming instead a Gaussian prior. For Fisher's reply, see Fisher (1921); see also Pearson (1968, pp. 452-454), Edwards (1974) and Box (1978, p. 79).
3. Egon Pearson (1968) gives the text of the letter from Pearson to Fisher rejecting the 1921 paper. In addition to the two papers just mentioned, Pearson had also earlier rejected a note by Fisher briefly criticizing an article in the May 1916 issue of *Biometrika*; see Pearson (1968, pp. 454-456). See also Pearson (1990).
4. "[H]e [Pearson] gained the devoted service of a number of able assistants, some of whom he did not treat particularly well. He was prolific in magnificent, or grandiose, schemes capable of realization perhaps by an army of industrious robots responsive to a magic wand. . . . The terrible weakness of his mathematical and scientific work flowed from his incapacity in self-criticism, and his unwillingness to admit the possibility that he had anything to learn from others, even in biology, of which he knew very little. His mathematics, consequently, though always vigorous, were usually clumsy, and often misleading. In controversy, to which he was much addicted, he constantly showed himself to be without a sense of justice. His immense personal output of writings . . . left an impressive literature. The biological world, for the most part, ignored it, for it was indeed both pretentious and erratic" (Fisher, 1956, pp. 2-3). As Savage has noted, Fisher "sometimes published insults that only a saint could entirely forgive" (Savage, 1976, p. 446). See also Kendall (1963, p. 3) and Barnard (1990, p. 26).
5. Karl Pearson's Laplacean view of probability is most carefully set out in Chapter 4 of his *Grammar of Science* (1892); see also Pearson (1920). Although little read today, the impact of Pearson's *Grammar* in his own time was considerable. (Mach's *Science of Mechanics*, for example, was dedicated to Pearson.) For the influence of the *Grammar* on Neyman, see Reid (1982, pp. 24-25).
6. I owe the material in this section to the generosity of Dr. A. W. F. Edwards, who has made available to me a considerable body of information he collected during the 1970s about the Rothamsted origins of the fiducial argument.
7. Fisher (1935c), in the chapter entitled "The Generalisation of Null Hypotheses. Fiducial Probability."
8. Because Fisher later distanced himself so emphatically from Neyman's viewpoint, the development of the theory of confidence intervals eventually came to be associated almost exclusively with Neyman (and his school). But, although Fisher disagreed with Neyman's behaviorist interpretation and unconditional uses of confidence intervals, Fisher's priority in the discovery of the method itself–in terms of publication relative to Neyman–seems largely unappreciated. [*Relative* rather than absolute priority: "E. L. Lehmann has pointed out that as far as computation (as opposed to logic) is concerned there is a long tradition of constructing confidence intervals involving Laplace and Poisson, followed by Lexis and one may add Cournot" (Hacking, 1990, p. 210). The reference is to a 1957 technical report written by Lehmann; in a footnote, Hacking notes that "Lehmann's paper has never been published, originally because he did not wish to offend Neyman" and cites as his source a personal letter from Lehmann dated 5 July 1988.]
9. Fisher does not specify with exactitude the necessary conditions on $F$, and his notation has the disadvantage that it does not distinguish between the random variate $T$ and an observed value of that variate.
10. Using this time the example of estimating $\sigma$ from $s$, on the basis of a random sample from a normal population, Fisher wrote:

    > Now we know that the inequality $[s > s_{0.01}(\sigma)]$ will be satisfied in just 1 per cent. of random trials, whence we may infer that the inequality $[\sigma < \sigma_{0.99}(s)]$ will also be satisfied with the same frequency. Now this is a probability statement about the unknown parameter $\sigma$. (Fisher, 1933, pp. 347-348)

11. This is underscored in a letter of Fisher to Fréchet several years later (26 February 1940; Bennett, 1990), where such a frequency is said to be "a definition of the phrase fiducial probability," one that Fisher had "no objection to regarding . . . as an arbitrary definition" (p. 127).
12. Fisher termed a probability resulting from the fiducial argument a "fiducial probability," to distinguish it from an "inverse probability" (Fisher, 1933, p. 83; 1945, p. 129), but stressed that while the terminology was intended to draw attention to the novel mode of derivation employed, such probabilities did not differ in kind from ordinary mathematical probabilities (Fisher, 1936) and that "the concept of probability involved is entirely identical with the classical probability of the early writers, such as Bayes" (Fisher, 1956, p. 54).

    For the distinction between probability and likelihood, see Fisher (1921, pp. 24-25; 1922, pp. 326-327).
13. "For a time this led me to think that there was a difference in logical content between probability statements derived by different methods of reasoning. There are in reality no grounds for any such distinction" (Fisher, 1956, p. 59).
14. For example, Wilks (1938).
15. As Fisher (1939b) later put it: "To all, I imagine, it [the fiducial argument] implies at least a valid test of significance expressible in terms of an unknown parameter, and capable of distinguishing, therefore those values for which the test is significant, from those for which it is not" (p. 384). See also Fisher (1935b, pp. 50-51).
16. "[Fiducial inferences] are certainly not statements of the distribution of a parameter $\theta$ over its possible values in a population defined by random samples selected to give a fixed estimate $T$" (Bennett, 1990, p. 124).
17. The fiducial distribution is the "*aggregate* of all such state-

ments as that made above" (Fisher, 1936, p. 253); "une loi de probabilité pour $\mu$ qui correspondra à l'ensemble des résultats trouvés plus haut" (CP 156, p. 155).

18. For example, in a letter in 1934 to Harold Jeffreys Fisher considered the amusing example of a traveller landing by parachute in a city and finding that he is one kilometer from its center. If the city is assumed circular and the position of the traveler random, then the fiducial probability that the radius of the city exceeds $R$ kilometers is $1/R^2$. Thus, the "fiducial median city has a radius $\sqrt{2}$ kilometres and an area $2\pi$. The fiducial mean radius is 2 km. and the fiducial mean area is infinite" (Bennett, 1990, pp. 160–161).
19. In particular, Fisher (1922, 1925, 1930, 1934, 1935b, 1936).
20. Fisher (1934, pp. 6–7; 1935b, p. 40).
21. Bennett (1990, p. 61).
22. Fisher (1934, p. 4).
23. Fisher (1922, p. 326); see also Bennett (1990, pp. 172–173). Fisher considers the theories of Ellis and Cournot to be "sound" (Bennett, 1990, p. 61).
24. Fisher (1925, p. 700; 1934, p. 7). "I myself feel no difficulty about the ratio of two quantities, both of which increase without limit, tending to a finite value, and think personally that this limiting ratio may be properly spoken of as the ratio of two infinite values when their mode of tending to infinity has been properly defined" (Bennett, 1990, p. 151; see also pp. 172–173).
25. CP 109, p. 7; CP 124, p. 81. Fisher thought that failure to maintain a clear distinction between the hypothetical and experimental value of probability was responsible for the lack of universal acceptance of the frequency theory (Bennett, 1990, p. 61).
26. This discussion is unfortunately omitted from Fisher's *Collected Papers*. All quotations in this section, unless otherwise stated, are from the report of the discussion at the end of Neyman (1934).
27. Strictly speaking, it would have been more accurate here for Fisher to have referred to the fiducial *argument*, rather than fiducial *probability*.
28. Note the use of the expression "fiducial statements based on inefficient estimates"; the fiducial argument may be employed in such cases, although care is needed in the use and interpretation of the resulting intervals. Fisher did not abandon this stand after his break with Neyman. In his 1956 book, Fisher wrote that "[confidence limits,] though they fall short in logical content of the limits found by the fiducial argument, and with which they have often been confused, do fulfil some of the desiderata of statistical inferences" (p. 69).
29. Fisher (1935a). In his introduction to the 1930 paper in *Contributions to Mathematical Statistics* (Fisher, 1950), Fisher states that "it should also have been emphasised that the information [supplied by a statistic employed in a statement of fiducial probability] as to the unknown parameter should be exhaustive" (pp. 392–393), and Fisher (1956) states that "though the [correlation coefficient] example was appropriate, my explanation left a good deal to be desired" (p. 57). See also Bennett (1990, pp. 81–82).
30. During the discussion, Neyman complimented Fisher on using a "remarkable device" to alter the problem "in such an ingenious way" (Fisher, 1935b, p. 76).
31. One reason for this particular interest might have been Fisher's often expressed view that the fiducial argument had not been noted earlier because of the "preoccupation" of earlier authors with discontinuous variates, to which the particular argument given in his 1930 paper did not apply (see, e.g., Fisher, 1935a, p. 391; 1941, p. 323).
32. Neyman later wrote, "Fisher took part in the discussion, and it was a great surprise to the author to find that, far from recognizing them as misunderstandings, [Fisher] considered fiducial probability and fiducial distributions as absolutely essential parts of his theory" (Neyman, 1941, p. 129). Although one cannot be sure how closely the published text of Fisher's remarks mirrors his actual words, Neyman's statement is certainly not supported by the published version of the discussion. Far from asserting that fiducial probabilities were a novel element of his theory, Fisher agreed with Neyman that they did not differ from ordinary probabilities, the adjective "fiducial" only being used to indicate "a probability inferred by the fiducial method of reasoning, then unfamiliar, and not by the classical method of *inverse* probability." The term "fiducial distribution" itself does not appear in Fisher's discussion. (Fisher did state that "with a single parameter, it could be shown that all the inferences might be summarized in a single probability distribution for that parameter, and that, for this reason, all were mutually consistent; but it had not yet been shown that when the parameters were more than one any such equivalent frequency distribution could be established.") Neyman would appear to be largely projecting back to 1935 ideas and statements made only later by Fisher.
33. It is interesting to contrast the treatment of this period in Constance Reid's biography of Neyman (Reid, 1982) and Joan Fisher Box's biography of Fisher (Box, 1978). Reid, who interviewed both Neyman and Pearson, paints a picture of continuing cordial relations until March 1935: "Jerzy – to start with – got on quite well with Fisher" (p. 114, quoting Egon Pearson); throughout the spring of 1934, "Neyman continued to be on good terms with Fisher; and he was invited, as he recalls, several times to Rothamsted" (p. 116); the next fall, "Neyman continued to receive friendly invitations to Rothamsted" (p. 120); and in December Neyman's "highly complimentary" remarks on Fisher's RSS paper (read on 18 December 1934) "drew grateful words from the beleaguered Fisher" (p. 121). In contrast, Box writes that after Fisher nominated Neyman for membership in the ISI in May 1934, "Neyman sniped at Fisher in his lectures and blew on the unquenched sparks of misunderstanding between the departments [of Genetics and Statistics at University College London] with apparent, if undeliberate, genius for making mischief," resulting in "open conflict" after the reading of Neyman's 1935 paper, whose "condescending attitude would have been galling, even if the conclusion had been sound" (p. 263).
34. "Professor R. A. Fisher, in opening the discussion, said he had hoped that Dr. Neyman's paper would be on a subject with which the author was fully acquainted, and on which he could speak with authority, as in the case of his address to the Society delivered last summer. Since seeing the paper, he had come to the conclusion that Dr. Neyman had been somewhat unwise in his choice of topics." First describing a statement in Neyman's paper as "extraordinary," Fisher later asked "how had Dr. Neyman been led by his symbolism to deceive himself on so simple a question?" and ended by referring to "the series of misunderstandings which [Neyman's] paper revealed" (Neyman, 1935a, pp. 154–157. Fisher's posthumously published *Collected Papers* presents a highly sanitized version of these comments.)
35. During the next several years, Fisher would provoke a series of needless professional and personal confrontations besides that with Neyman: lashing out at this old friend Gosset a year before the latter's death, exacerbating a long controversy in population genetics with Sewall Wright and worsening relations with his wife, which led to permanent separation. Fisher had moved in 1933 from the congenial atmosphere of Rothamsted to a divisive University College London, and it is possible that in this perhaps unfortunate change lies much

of the explanation; another important factor may have been the decidedly unfriendly reception given to his 1935 Royal Statistical Society paper.

36. Although Neyman's comments largely suggest a personal attack, the wording "my ideas" and "those with whom I was working" might also be taken to include attacks on Neyman's work not directly naming him, and attacks on others, not necessarily coauthors. I have not found many instances of the former, however, and the disputes with Bartlett (discussed below), Wilson and Barnard do not appear to fall into the category delineated by Neyman. Who Neyman might have had in mind is unclear.
37. This statement refers only to direct exchanges between the two, and not to others who may have served as proxies; see, for example, the paper by Yates (1939) and its discussion by Neyman (1941).
38. The paper testifies to the sudden deterioration in relations between Fisher and Neyman. Where just a few months earlier Fisher had referred approvingly to Neyman's 1934 paper (see note above), now Fisher wrote, "Dr. J. Neyman has unfortunately attempted to develop the argument of fiducial probability in a way which ignores the results from the theory of estimation, in the light of which it was originally put forward. His proofs, therefore, purport to establish the validity of a host of probability statements many of which are mutually inconsistent" (Fisher, 1935a, p. 319).
39. The statement being made is in fact quite strong. The phrase "it appears that" does not intend the qualified assertion "it *would seem* that" but the unqualified assertion "it *is seen* that"; compare Fisher's use of the expression in the preceding paragraph. When the paper was reprinted (in his *Collected Papers*), Fisher had added the curious footnote, "After appears, insert likely"!
40. If, given $\alpha$ and $x$, $\alpha_1(x)$ and $\alpha_2(x)$ denote the unique numbers satisfying the dual constraints $\alpha_1(x) - \alpha_2(x) = \alpha$ and $x + c_{\alpha_2}(x) = -(x + c_{\alpha_1}(x))$, then it is *not* the case that $P_\mu[X + c_{\alpha_2}(X) < \mu < X + c_{\alpha_1}(X)] = \alpha$ (see, e.g., Pedersen, 1978).
41. "[One source of paradoxes] is the introduction, into an argument of this type, of fixed values for the parameters, an introduction which is bound to conflict with the fiducial distributions derivable from the data" (Fisher, 1937, p. 370). Why this would conflict with the fiducial distribution Fisher did not state. In later papers, Fisher attempted to deal with this difficulty by dogmatic decree. Thus, "The notion of repeated sampling from a fixed population has completed its usefulness when the simultaneous distribution of $t_1$ and $t_2$ has been obtained" (Fisher, 1941, p. 148).
42. In his introduction to this paper in *Contributions to Mathematical Statistics* (Fisher, 1950), Fisher admitted as much when he pointed to "the first section, in which the logic of the [Behrens-Fisher] test is discussed" and noted that "the principles *brought to light* seem essential to the theory of tests of significance in general" [emphasis added]. But given that those principles had been "brought to light" in 1939, his charge – in the very same sentence – that they had "been most unwarrantedly ignored" by Neyman and Pearson in a paper written *seven years earlier* is curious to say the least.

An important factor contributing to this shift may have been Fisher's rereading of Gosset's papers while drafting an obituary notice for "Student" (Fisher, 1939c). Fisher's method, first given there, for estimating the median or other percentiles of a distribution "irrespective of the form of curve" (pp. 4–5) stands in striking contrast to his earlier criticism of a paper by Harold Jeffreys (Jeffreys, 1932; Fisher, 1933). Jeffreys had asserted that given two observations $x_1$ and $x_2$, the probability is 1/3 that a third observation $x_3$ will lie between the first two; Fisher now asserted (p. 4) that given two observations $x_1$ and $x_2$, the probability is 1/2 that the median lies between them. But the repeated sampling argument that Fisher had employed to ridicule Jeffrey's statement in 1933 could be easily modified to attack his own assertion in 1939. (Note also that Fisher's new emphasis on the uniqueness of the sample at hand was also justified by pointing to the wording used earlier by Student; see Fisher, 1939a, p. 175).

Another important factor contributing to Fisher's shift in viewpoint was undoubtedly his lengthy exchange of letters with Harold Jeffreys between 1937 and 1942 (Bennett, 1990, pp. 161–178). See especially Fisher's comment that "I have just reread your note on the Behrens-Fisher formula.... I think your paper enables me to appreciate your point of view [i.e., conditional and Bayesian] a great deal better than I have previously done" (pp. 175–176).

43. Indeed, initially Fisher does not seem to have recognized the inconsistency of the two positions: in a letter to Fréchet in 1940 (26 January; Bennett, 1990, p. 121) Fisher reiterated the position of his 1930 paper that in a statement of fiducial probability the statistics involved are not considered as fixed and that such a statement differs in logical content from one of inverse probability. (See also Fisher's letter to Fréchet dated 10 February 1940; Bennett, 1990, p. 124.) The clash with the language in a paper of only five years later (Fisher, 1945, quoted below in endnote 48) is particularly striking.
44. Fisher pointed out that a distribution is not determined, as had been claimed, by its moments; see Reid (1982, p. 103). Ironically, Fisher may have served as one of the referees for the paper when it was submitted to the *Philosophical Transactions*, reporting favorably on it; see Reid (1982, pp. 102–104).
45. "With samples of more than 2, I should expect some differences fiducially significant to be found insignificant, if tested for some particular values of the variance ratio, these being ratios which the data themselves had shown to be unlikely" (Fisher, 1937, p. 375). In his 1937 paper, Fisher mathematically demonstrated the conservative nature of the Behrens-Fisher test for samples of size two, presumably to verify the universal validity of the phenomenon noted by Bartlett in several specific instances. Although numerical studies suggest that the Behrens-Fisher test is indeed conservative for all sample sizes, as Fisher later asserted in 1950, a mathematical demonstration of this fact is still lacking today (see, e.g., Wallace, 1980, p. 137)!
46. In some cases, of course, Fisher may simply have come to believe in a new position with such force and conviction that he simply forgot that there had ever been a time when he thought otherwise. For example, in a letter to Barnard in 1954 Fisher criticized Neyman for ignoring "my warning [in Fisher, 1930] that the fiducial distribution would be invalid to any one possessing knowledge *a priori* in addition to the observed sample" (Bennett, 1990, pp. 9–10). In reality, as we have seen, far from having issued such a warning, Fisher clearly takes the opposite position!
47. Note the contrast with Fisher's statement in 1935 that "to attempt to define a prior distribution of $\mu$ which shall make the inverse statements coincide numerically with the fiducial statements is really to slur over this distinction between the meaning of statements of these two kinds" (Fisher, 1935a, p. 392).
48. This paper, although it does not refer to Neyman (1941), was clearly intended as a reply to it. ["The purpose of this note is therefore to discuss . . . the process of reasoning by which we may pass, without arbitrariness or ambiguity, from forms of statement in which observations are regarded as random variables, having distribution functions involving certain

fixed but unknown parameters, to forms of statement in which the observations constitute fixed data, and frequency distributions are found for the unknown parameters regarded as random variables" (Fisher, 1945, p. 507).]

49. See, for example, the rather jaundiced view of Raymond Birge, the Berkeley physicist, quoted in Reid (1982, p. 144). Savage's assessment was much more sympathetic (and probably more accurate): "I am surely not alone in having suspected that some of Fisher's major views were adopted simply to avoid agreeing with his opponents. One of the most valuable lessons of my rereading is the conclusion that while conflict may have sometimes somewhat distorted Fisher's presentation of his views, the views themselves display a steady and coherent development" (Savage, 1976, p. 446, references omitted).
50. "From the time I first introduced the work, I have used the term fiducial probability rather strictly, in accordance with the basic ideas of the theory of estimation. Several other writers have preferred to use it in a wider application, without the reservations which I think are appropriate" (Fisher, 1939b, p. 384).
51. Creasy's paper (1954) dealt with the problem of assigning fiducial limits to a ratio of normally distributed means (the so-called Fieller-Creasy problem); see Wallace (1980). It is generally agreed, even by many of Fisher's most sympathetic readers, that he was unfair in his critical response to Creasy's paper; see, for example, Box (1978, p. 459) and Wallace (1980, p. 141).
52. The more general case of estimating the percentiles of a distribution on the basis of a sample of size $n$ had been discussed earlier by Fisher in his obituary of 'Student' (Fisher, 1939c).
53. Fisher's correspondence with Fréchet in 1940 (Bennett, 1990, pp. 118–134) is particularly interesting in this regard. Repeatedly pressed by Fréchet to justify the transition, Fisher eventually argued that the particular sample (and therefore the resulting interval) could be regarded as "an event drawn at random from the population investigated" and therefore that the single-case probability of coverage could be identified with the frequency of coverage in the population as a whole.
54. See Fisher's remarks in the preface to the 13th edition of *Statistical Methods for Research Workers* (Fisher, 1925, 13th ed, 1958).
55. The language of this passage suggests that, as late as the beginning of 1955, Fisher had not yet arrived at his recognizable subset justification for the fiducial argument.
56. See, for example, Box (1978, pp. 441–442). I. J. Good reports that he had been told Fisher liked his 1950 book *Probability and the Weighing of Evidence* (Savage, 1976, p. 492; see also Bennett, 1990, p. 137). As Barnard notes, Jeffreys "got on extremely well with Fisher" (Barnard, 1990, p. 27), as is evident also from their published correspondence.
57. See also Fisher (1958, 1959, 1960). In a letter to D. J. Finney dated 15 March 1955, Fisher says he has "recently been thinking a little about the semantics" of the word "probability" (Bennett, 1990, p. 96).
58. Fisher's distaste for and suspicion of axiomatics is evident throughout his published correspondence; see, for example, Bennett (1990, pp. 128–129, 175, 185, and 331).
59. For Fisher's exchange with Jeffreys, see Lane (1980). (In addition to the papers discussed by Lane, there is also an exchange between Fisher and Jeffreys that occurs at the end of Fisher's 1935 *Journal of the Royal Statistical Society* paper.)