International Statistical Review (2007), **75**, 2, 131–149 doi:10.1111/j.1751-5823.2007.00020.x

Information and Economics in Fisher's Design of Experiments

John Aldrich

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

Summary

In 1951 R. A. Fisher described what had been achieved in the 20th century so far: "we have learnt (i) To conserve in its statistical reduction the scientific information latent in any body of observations. (ii) To conduct experimental and observational inquiries so as to maximise the information obtained for a given expenditure." This paper asks what Fisher meant and, in particular, how he saw his work on experimental design as contributing to the objective of maximizing information for a given expenditure. The material examined ranges from detailed work on issues like "the information lost in measurement of error" to polemics against decision theory.

Key words: R.A. Fisher; design of experiments; information; decision theory.

1 Introduction

The themes usually associated with Ronald Fisher's work on experimental design are replication, randomization and local control, *not* information and economics; see e.g. Atkinson & Bailey (2001, p. 57). Yet consider Fisher's (1951, p. 54) summary of what statistics had achieved in the first half of the 20th century:

> we have learnt (i) To conserve in its statistical reduction the scientific information latent in any body of observations. (ii) To conduct experimental and observational inquiries so as to maximise the information obtained for a given expenditure.

In this formula, spanning what statisticians do with observations and how they choose them, Fisher is clearly referring to his own work—in (i) to the theory of estimation (1922, 1925b, 1934), and in (ii) to *The Design of Experiments* (1935), hereafter *Experiments* for short, and publications going back to 1925.

On another occasion Fisher (1947, p. 435) presented the "development of the theory of experimental design" in terms of the balance between (i) and (ii):

> [The finding that the amount of information extracted in the process of estimation could never exceed the quantity supplied by the data], combined with the practical fact that directly available processes of computation would extract almost always a very large fraction of the total available shifted the *moral balance*. . . . The weight of [the statistician's] responsibility

was thrown back on to the process by which the data had come into existence. So armed with amount of information as a practical tool, statisticians came to study what forms of experiment or what types of observational programs would yield the most information for a given expenditure in time, money and labor.

Was there anything behind this story and the spanning formula, or were they the froth celebrations call for? The standard developments, expositions, appreciations and critiques of Fisher's design work do *not* treat replication, randomization and local control as expressions of more fundamental concerns with information and economics because, presumably, the authors believed those ideas played no real part in the work or were better dropped; see for example, Cochran & Cox (1950), Youden (1951), Finney (1955), Yates (1964), Pearce (1979), Preece (1990) or Street (1990). I will argue that Fisher made a genuine effort to found his design work on those ideas and it is his failure which makes the formula and story sound hollow.

Fisher first published on information in 1922 and, according to Box (1978, p. 156), designed his first experiment in 1924. His early estimation work is described in Section 2 and his early design work in Section 3–6. How he brought information and design together is discussed in Section 7–9. Section 10 picks up the economic theme and compares Fisher's work with Wald's decision theory, which he knew well enough to attack, and with comparison-of-experiments theory, which he did not know at all, and which appears not to have known him. The concluding remarks in Section 11 include some speculations about why these themes have not been visible.

2 The Total Available Relevant Information

The theme of information and its conservation distinguish Fisher's estimation from earlier work in the theory of errors and biometry. Today "Fisher information" is either a heuristic device or a technicality in the asymptotic theory of maximum likelihood, but the intuitions and the asymptotic theory are survivals from a larger scheme. We need to notice the scheme to make sense of the "conservation" slogan and to identify the ideas Fisher applied to design. Only the basic ideas are needed for Fisher did not apply the more sophisticated concepts, such as ancillarity and second-order efficiency, to design. For more on the origins of the theory, see Aldrich (1997) and, for more on the later, more sophisticated theory, see Hinkley (1980, 1980a).

The theme of "conserving scientific information" first appears in the "Mathematical Foundations of Theoretical Statistics" (1922). However, the theory of information was more thoroughly developed in the "Theory of Statistical Estimation" (1925b). There (p. 709) the 'intrinsic accuracy' of the frequency curve f "as a means of estimating θ" is defined as the mean of $-\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)$ which "may equally be conceived as the amount of information in a single observation belonging to such a distribution".

The "efficiency" with which information is "isolated" (or "extracted" or "conserved") was also formalized. The first formulation, Fisher (1922, p. 316; 1925a, p. 13), was restrictive but the second (1925b, p. 714) had "the advantage of applying to finite samples and to other cases where the distribution is not normal". Fisher would find more to say about conserving information but the chief practical lesson was already there in 1922: don't use the method of moments (exposed as inefficient in Section 8–11), use maximum likelihood! Maximum likelihood underwrote the "practical fact that directly available processes of computation would extract almost always a very large fraction of the total available".

The 1922 paper supplied tools for the theory of estimation but it also cast the problem of Statistics as the problem of estimation and Statistics as "the study of methods of the reduction of data" went unchanged into the *Statistical Methods for Research Workers*, or *Methods* for short. This is *not* today's "descriptive statistics" for Fisher's (1925a, pp. 1, 7) statement, "It is the object of the statistical processes employed in the reduction of data to exclude ... irrelevant information, and to isolate the whole of the relevant information contained in the data", is to be understood in terms of sufficient statistics and parameters of "hypothetical infinite populations."

At mid-century this definition is still there as lesson (i), redrawn to make way for (ii). Lesson (i) was never a good description of what Fisher's non-experimental statistics was about. The information objective of the first chapter of the *Methods* does not fit the tests of significance that fill the rest. In the first chapter and in the "Mathematical Foundations" (p. 304) testing is confined to checking the validity of the specification (the "mathematical form of the population"). Fisher's ideas on testing and on information were broadly compatible in that the test statistics were usually based on quantities with good information credentials but he never found it necessary to make really tight connections. See Section 8–9 below.

There was more of a journey for (ii) and, by one test, it never arrived! Fisher (1925a, p. 9) stated that, when the consequences of a hypothesis agree with observations, it is "justified at least until fresh and more stringent observations are available." The discussion of experiments later in the book made that passivity obsolete on publication. After 1935 an advertisement was added, "The author has attempted a fuller examination of the logic of planned experiments in his book, *The Design of Experiments*." Point (ii) materialized in the 1930 edition of the *Methods* (see Section 7) in the last chapter. It never reached the first and statistics remained "the study of methods of the reduction of data."

To summarise, by 1922 Fisher had a vision, or at least a slogan, and the basis for believing that "directly available processes of computation would extract almost always a very large fraction of the total available". How statisticians interested in design came to be "armed with amount of information as a practical tool" is a more involved story.

3 Decreasing the Standard Deviation

From his first published work on design Fisher emphasised "the two *desiderata* of the *reduction of error* and of the *valid estimation* of error", as he put it in 1926 (p. 508). The "valid" estimation of error—and randomization as a way of guaranteeing it—was a new concern but the reduction of error was not; it was already seen as the main goal in designing experiments.

From joining Rothamsted in 1919 Fisher was in continuous contact with W.S. Gosset ('Student') and from 1923 the design of experiments figured in their letters (McMullen, 1970); see Pearson's '*Student*' (1990) for an account of 'Student's work and his interaction with Fisher. Although Gosset had only one pre-war publication on design, he had been giving advice on the statistical aspects of experiments from 1904 and, insofar as there was a statistical authority on the design of experiments, he was *it*. The publication was the appendix to Mercer & Hall's "The Experimental Error of Field Trials" (1911), an important paper on the variability of yields in uniformity trials.

The appendix gave "a method of arranging plots so as to utilize a given area of land to the best advantage in testing two varieties." Like others later, including Fisher (1925a) and Engledow (1925), 'Student' used data from the paper to illustrate his reasoning. He imagined a field divided into plots sown "chequer fashion" with the two kinds of seed. The objective is to "decrease the standard deviation and to increase our certainty." On the size of the plot and the advantage of comparing adjacent plots, he concludes

> Roughly speaking, one-twentieth acre plots of mangolds would require at least twice as much land as one-two-hundredth acre plots in order that we may place as much confidence in the results. . . .
> Hence it is clearly of advantage to use the smallest practicable size of plot.
> Also the advantage of comparing adjacent plots is apparent in these examples since with roots less than two-thirds of the land is required to give the same accuracy as random comparison. . .

"Confidence" and "accuracy" are both determined by the size of the standard deviation of the difference between adjacent plots. In 1923 'Student' entertained the possibility of using calculus to determine the size of plot to give the "minimum probable error per acre" but considered the data inadequate (p. 279). Fisher retained the goal of decreasing the standard deviation but re-phrased it as increasing the information, although the rephrasing took rather a long time to effect.

4 "Modern Methods" and the Analysis of Variance

Fisher's ideas on design appear fully-formed in the *Methods* (1925a) as the "principles underlying modern methods of arranging field experiments". "Modern methods" was an extraordinary description for methods that hitherto had existed only in Fisher's head! The principles appear in the "further applications" of the analysis of variance for Fisher (1925a, p. 224) held that, "The statistical procedure of the analysis of variance is essential to the understanding of [those] principles." Because the analysis of variance provided the framework for all of Fisher's work on agricultural experiments, it requires examination.

Fisher's best explanation of the assumptions underlying the analysis of variance is an explanation he wrote for Gosset and which appears as a footnote in 'Student' (1923).

The passage in 'Student' (1923, p. 283n) begins:

> The yield obtained in any experiment is the sum of three quantities, one depending only on the variety; a second depending only on the 'trial'; and a third, which may be regarded as the 'experimental error' varying independently of variety and trial in a normal distribution with a standard deviation which it is desired to estimate.

In Fisher's notation the model explaining the yields X_{pq} is thus (the error symbol is mine)

$$X_{pq} = A_p + B_q + \varepsilon_{pq}, \qquad \varepsilon_{pq} \sim IN(0, \sigma^2), \qquad p = 1, \ldots, m; q = 1, \ldots, n.$$

Fisher estimates the quantities A_p, B_q by least squares, minimizing

$$\sum (X_{pq} - A_p - B_q)^2. \tag{1}$$

Evidently (1) will be a minimum if

$$A_p + B_q = \overline{X}_p + \overline{X}_q - \overline{\overline{X}}$$

where $\overline{X}_p$ is the mean of the values obtained with variety p, $\overline{X}_q$ the mean of the values obtained with trial q, and $\overline{\overline{X}}$ is the general mean." Next comes the analysis of variance table:

The actual evaluation is most conveniently carried out in the following form of the analysis of variance:

Variance	Degrees of freedom	Sum of squares
(a) Due to variety	$m-1$	$n\sum_1^m(\overline{X}_p-\overline{\overline{X}})^2$
(b) Due to trial	$n-1$	$m\sum_1^n(\overline{X}_q-\overline{\overline{X}})^2$
(c) Random variation	$(m-1)(n-1)$	$\sum_1^m\sum_1^n(X_{pq}-\overline{X}_p-\overline{X}_q+\overline{\overline{X}})^2$
(d) Total	$mn-1$	$\sum_1^m\sum_1^n(X-\overline{\overline{X}})^2$

The sum of squares in line (c) being calculated by subtracting the values of lines (a) and (b) from the total. If either variety or 'trial' were without significant effect on the yield, the corresponding mean square would not differ significantly from that of line (c). To test the significance of such a difference we may use the fact that the estimates of variance in (a), (b) and (c) are all independent, and when m and n are fairly large the natural logarithms of the mean square has standard deviation $\sqrt{2/n_1}$ where n_1 is the number of degrees of freedom.

Here is the familiar table with the variance and degrees of freedom identities and also the idea of testing whether distinct estimates of a variance are significantly different. The normal approximations were soon replaced by the exact $z\ (=\frac{1}{2}\ln F)$ distribution. Fisher introduced z in his (1924) "On a distribution yielding the error functions of several well known statistics" and used it in Ex. 41 of Fisher (1925a, p. 203).

When Fisher took this framework from *analysis* and applied it to *design* he introduced randomization at the same time. The next section considers the status of normal theory when there is randomization. This may seem a detour from our quest for information in design but the status of normal theory matters for Fisher knew how to do normal theory information calculations, but he never defined information in nonparametric situations.

5 Randomization and the Theory of Errors

Fisher went on using the normal theory methods of 1923—with the z-refinement—after he adopted randomization. Numerous authors, including Welch (1937) and Kempthorne (1952), have investigated the relationship between randomization and normal theory, sometimes justifying Fisher's practice and sometimes not. Naturally, commentators on Fisher have asked what was in *his* mind when, for example, he wrote

> The estimate of error is valid, because, if we imagine a large number of different results obtained by different random arrangements, the ratio of the real to the estimated error, calculated afresh for each of these arrangements, will be actually distributed in the theoretical distribution by which the significance of the result is to be tested. (Fisher, 1926, p. 507)

Kempthorne (1966) and Box (1980) argue that Fisher considered the normal theory analysis an approximation to an exact test based on the randomization distribution, while Hinkley (1980b, p. 583) appeals to several authorities, including Yates, to support the view, "the purpose of randomization in the design of agricultural field experiments was to help ensure the validity of normal-theory analysis". I think the latter better represents Fisher's position—whether, or not, it is defensible!

The statement at the beginning of chapter X of the *Experiments* seems clear:

> The foregoing chapters, III to IX, have been devoted to cases to which the theory of errors is appropriate . . . In this kind of hypothesis all discrepancies classified as error, and not eliminated by equalisation or regression, are due to variation in the material examined, following the normal law of errors with a definite and constant, but unknown variance. (Fisher, 1935, p. 187)

In support of their view Kempthorne and Box quote a well-known passage from earlier in the book, Fisher (1935, p. 51),

> the physical act of randomisation, which . . . is necessary for the validity of any test of significance, affords the means, in respect of any particular body of data, of examining the wider hypothesis in which no normality of distribution is implied.

The "wider hypothesis" was not something that required investigation because of randomization. It considered the possibility that the "variation in the material examined"—the ε_{pq} of the last section—may be non-normal. While Fisher liked the ingenuity of his randomization test, he insisted that non-normality was not a genuine issue. Fisher (1926, p. 505) held that the "assumptions about the nature of field errors" used in the theory of errors have been "extensively verified in the examination of the results of uniformity trials." There are similar statements in Fisher (1929, 1935, p. 50) on the randomization test.

6 "Accuracy", "Precision" and "Information"

Fisher's (1925a, pp. 224–229) exposition of the "technique of plot experimentation" demonstrates two devices: randomization to produce a valid estimate of error and the use of randomized blocks to reduce error. Fisher rejected any "pre-arranged system" of choosing plots, going against what 'Student' had recommended in 1911 and 1923.

Fisher makes his points, not by a theoretical analysis, but by, what was, in effect, a single replication Monte Carlo experiment inspired by the real conditions of Mercer & Hall's (1911) uniformity trials. I will describe this in detail for it was the canonical example through which Fisher expressed his meaning in 1925(a,b) and re-expressed it in 1938a. To the 20 strips of land Fisher assigns five different treatments each in quadruplicate to them. The allocations are "found by shuffling 20 cards thoroughly and setting them out in order." The letters are just labels attached to plots for there is no difference in treatment. The one-way analysis of variance table is

Variance due to	D. F	Sum of Squares	Mean Square	St. Dev.
Treatment	4	58,725	14,681	121.1
Experimental error	15	231,041	15,403	124.1
	19	289,766	15,251	123.5

Fisher comments, "It will be seen that the standard error of a single plot estimated from such an arrangement is 124.1, whereas, in this case, we know its true value to be 123.5." The point is that the standard error of a single plot is close to the true standard deviation of this finite population of 20 numbers.

He next shows how, by blocking, the "accuracy of our comparisons is much improved; in fact the remaining variance is reduced to almost 55 per cent of its previous value":

> we may divide the 20 strips into 5 blocks, and impose the condition that each treatment shall occur once in each block; we shall then be able to separate the variance into three parts representing (i.) local differences between blocks, (ii.) differences due to treatment, (iii.) experimental errors; and if the five treatments are arranged at random within each block, our estimate of experimental error will be an unbiased estimate of the actual errors in the differences due to treatment. As an example of a random arrangement subject to the above restriction, the following was obtained:
>
> A E C D B | C B E D A | A D E B C | C E B A D.
>
> Analysing out, with the same data as before, the contributions of local differences between blocks, and of treatment, we find

Variance due to	D. F.	Sum of Squares	Mean Square	St. Dev.
Local differences	3	154,483	51,494	..
Treatment	4	40,859	10,215	..
Experimental error	12	94,424	7,869	88.7
Treatment + error	16	135,283	8,455	92.0

The value of "almost 55 per cent" for the reduction is obtained from the 8,455 of this table and the 15,251 of the previous one. The criterion of "improvement" is the same as Gosset's (Section 3). Fisher's "The Arrangement of Field Experiments" (1926) contains a similar comparison of a fully randomized scheme and a randomized block scheme, linking the two with the statement "On most land, however, we shall obtain a smaller standard error, and consequently a more valuable experiment, if we proceed otherwise." (p. 509)

From Fisher's words quoted in the Introduction—"armed with amount of information as a practical tool, statisticians came to study . . ."—it may be surprising that information was only fully installed in Fisher's writing in 1935. In the meantime, for example in Eden & Fisher (1927), "accuracy" and "precision" were used as glosses on the standard deviation. Yet information seems to have been in waiting. When Fisher & Eden (1929, p. 205) introduced a "precision index", they denoted it by I; it was calculated according to the formula

$$I^{-1} = 100\left(\frac{\sigma}{m}\right)^2,$$

where "σ is the standard deviation of the mean yield of each treatment, and m is the mean yield of all." The index was presented in the *Experiments* in Section 60, *Precision regarded as amount of information* (1935, p. 180) and I was identified as the "quantity of information". As such, it went into the *Methods* (7th edition, 1938).

Fisher (1938, Section 48, p. 275) applies the 1929/1935 information formula in a reconsideration of the 1925 example from the viewpoint of the "amounts of information elicited by different methods",

> If we take as having unit value an experiment giving comparable yields subject to a standard error of 10 per cent., the value of such an experiment as this, in quadruplicate, may be found by squaring one-tenth of the mean yield [$3285.75 = m$], multiplying by four (giving 431816)

and dividing by the mean square obtained by each method of procedure. For randomisation without blocks we have then

$$\frac{431816}{15250.8} = 28.18$$

units of information. Using randomised blocks we have 51.07 . . .

	Degrees of freedom for error	units of information
Randomisation of 20 plots	15	28.18
Randomisation in 4 blocks	12	51.07

In 1938 there was also an adjustment for position within the blocks based on the analysis of covariance but, as it does not involve any new principles, it is omitted from the table. There was also a refinement in evaluating these arrangements, which *did* involve a new principle. The refinement, introduced in the *Experiments* (1935), is considered in Section 9. But first we consider a type of experiment where the aim is not the detection of differences between treatments but the estimation of a parameter. Here the information objective in design was first made patent.

7 Information: "Design and Precision"

Besides agricultural experiments, based on the (normal) theory of errors, the *Experiments* treated two other kinds of experiment. The lady tasting tea (ch. II) was a non-parametric one-off but the other discussed in chapter XI represented a substantial body of work integrated into the theory of estimation. The examples in Fisher (1935, pp. 218–244) involve discrete distributions and the design interest ranges from the choice of sample size, when estimating the probability of success in Bernoulli trials, to the choice of schemes for estimating linkage in man. There was no randomization, as though we might reasonably "act in the faith that Nature has done the randomization for us" to use Fisher's words from 1947 (p. 436). This information-based design theory first appeared in 1930 when a new subsection was added to the estimation chapter of the *Methods*; it extended the running example of estimating linkage in plants to the choice of crosses to make.

Originally (see Section 2), Fisher used information for comparing the sampling distributions of estimators based on the same data and the idea of 1930 was that it could be used for comparing distributions arising from different experiments, provided that the same parameters were involved. There was something like this in the "Mathematical Foundations." Figure 1 of Fisher (1922, p. 322) displays a Pearson Type VII curve—the Cauchy distribution—and a normal curve of equal intrinsic accuracy. The point of the pairing is that for any (troublesome) Cauchy there is a (tractable) normal that is as difficult to locate; here intrinsic accuracy is interpreted as a measure of the difficulty a particular distribution poses for estimation.

"The Amount of Information: Design and Precision," Section 57.2 of Fisher (1930), compares the information to be gained from the first generation of plants, represented by (A), and that from breeding the second generation from different kinds of plants, represented by (B) or (C). The interest is in a recombination fraction p and Fisher is comparing three ways of getting information about p. The three experiments amount to sets of Bernoulli trials, where the probability of a success is related to the parameter p in three different ways. As usual, the argument unfolds through a numerical example based on actual data. Having estimated the recombination parameter, Fisher gives values for the additional information per plant under the three schemes: "At 6.345 per cent recombination [the estimated value] the numerical contributions under (A), (B) and (C) are .006051, 31.77 and 35.58. The second year's classifications thus give nearly 5000 and 6000 times as much information per plant as the first year's classification." (Fisher, 1930, p. 267).

The information/resources theme, that is, point (ii) from the mid-century (Introduction), appears when Fisher (1930, p. 268) describes the value of such calculations:

> The advantage of examining the amount of information gained at each stage of the experiment lies in the fact that the precision attainable in the majority of experiments is limited by the amount of land, labour and supervision available, and much guidance may be gained as to how these resources should best be allocated, by considering the quantity of information to be anticipated. In the experiment in question, for example, it appears that progenies from **OhAr** plants [(C)] are somewhat more profitable than those from **ohAr** plants [(B)].

Fisher's use of the estimated value in computing the expected information dramatizes the comparisons but is only applicable in situations where the implications of the results of a pilot experiment are being considered. The examples in chapter XI of the *Experiments* are all set at the pre-data stage of the investigation. One arrangement is better than another, although by how much will depend upon the (unknown) value of the parameter.

"Design and Precision" the sub-title of Section 57.2 indicates a second theme, the separation of points (i) and (ii) in a way that resembles Raiffa & Schlaifer's (1961) posterior and preposterior analysis:

> If . . . our object is merely to assign a standard error to a particular result, we may estimate the amount of information available directly by differentiating the expression for $\partial L/\partial p$ in the equation for maximum likelihood, using the actual numbers recorded in the classes observed. . . [This] gives 3725 as the total amount of information upon which our estimate has been based and 1.638 as the standard error of the estimate of the recombination percentage. It should be noted that an estimate obtained thus is in no way inferior to one obtained from the theoretical expectations; only that it gives no guidance as to the improvement of the conduct of the experiment. It might be said that owing to chance, . . ., the experiment has given the small amount of 11 units less information than should be expected [3736] from the numbers classified.

In the 4th edition Fisher (1932, p. 193) re-did the calculations and changed his conclusion about the informativeness of this experiment, compared to expectations. However, this did not affect the general point.

In 1930 and 1935 Fisher showed by example how to use information in design. He did not provide a general account of where the principle of using expected information in choosing between experiments would lead, or the obstacles it might face, for example if there is no experiment that produces more information in all parts of the parameter space. By the standards of later comparison-of-experiments theory—see Section 10 below—this is very primitive work.

1930 saw the introduction of the theme of information as a guide to the best allocation of resources in experiments. By 1933 the arithmetic on which the "shifting moral balance" was based had been done:

> By applying statistical methods not only to the interpretation but also to the design of experiments it is not uncommon for the value of the experiment to be increased five or tenfold, a result which could not be obtained from improved methods of interpretation only, unless previous methods had been excessively inefficient. (Fisher, 1933, p. 50; cf. 1935, p. 217)

Here the commensurability of the "value" of an experiment and the "efficiency" of a "method of interpretation" is taken for granted.

Perhaps "methods of interpretation" was meant to suggest something more comprehensive than estimation. Fisher's agricultural experiments usually led to a test but the lesson of 1930 was how better design produces better estimates—on average at least. The complicated story of Fisher's attempt to use information to show how better design produces better tests is told in the next two sections. Here it is much less clear what he was trying to do, or, what he thought he had achieved! By contrast, the link between design and testing in the Neyman–Pearson theory through power is clear. Neyman *et al.* (1935) describe that link and, although Fisher opened—very aggressively—the discussion of this paper, he seems to have registered this aspect of the paper only 20 years later. See Section 9.

8 Precision and Degrees of Freedom

In Section 6 we saw the arithmetic of blocking to reduce the standard deviation as given in Chapter VIII of Fisher (1925a). However, there is also a discussion of blocking—and replications—in Chapter V, on the comparison of means, where the design issues are linked more clearly to obtaining a better test and a tighter interval statement. The discussion and examples range across observational and experimental situations without distinction.

Examples 17 and 18 of Fisher (1925a, pp. 103–105) present 95% intervals for the difference between men's and women's heights. Fisher comments on the first unpaired comparison,

> we have treated the two samples as *independent*, as though they had been given by different authorities; as a matter of fact, in many cases brothers and sisters appeared in the two groups; since brothers and sisters tend to be alike in stature, we have overestimated the probable error of our estimate of the sex difference. Wherever possible, advantage should be taken of such facts in designing experiments. In the common phrase, sisters provide a better "control" for their brothers than do unrelated women. The sex difference could therefore be more accurately estimated from the comparison of each brother with his own sister. (p. 104)

The first interval was from $4\frac{1}{2}$ to 5 inches but with the paired comparison (Example 18), "we may estimate the mean sex difference as $4\frac{3}{4}$ to 5 inches". The calculations are based on large-sample normality and Fisher presented no interval inferences using Student's distribution, which is the main topic for this chapter; for more on interval inference in Fisher's work, see Aldrich (2000).

Having introduced Student's distribution, Fisher (1925a, p. 111) discussed the impact of changing the number of replications:

> The use of "Student's" distribution enables us to appreciate the value of observing a sufficient number of parallel cases; their value lies not only in the fact that the standard error of a mean decreases inversely as the square root of the number of parallels, but in the fact that the

> accuracy of our estimate of the standard error increases simultaneously. The need for duplicate experiments is sufficiently widely realised; it is not so widely understood that in some cases, when it is desired to place a high degree of confidence (say $P = .01$) on the results, triplicate experiments will enable us to detect with confidence differences as small as one-seventh of those which, with a duplicate experiment, would justify the same degree of confidence.

The figure of "one-seventh" reflects the .01 points on t with 1 degree of freedom (63.657) and with 2 (9.925). In Neyman–Pearson terms, Fisher is *not* describing how the power function changes but how the .01 critical region of the t-test changes shrinks as the sample size changes. Fisher emphasised the importance of increased numbers of degrees of freedom elsewhere too, for example in Fisher & Wishart (1930, p. 12)

> A valid test could be made based on only one degree of freedom for error, but as the table for z shows the values of z exceeded in 5 per cent. or 1 per cent. of trials are very large, and in consequence relatively large real effects must be judged insignificant, or in other words, escape detection.

Examples 19–20 of the *Methods* continue the paired versus unpaired comparison using the Cushny and Peebles data on the reaction of 10 patients to two drugs. This gave Fisher a second example showing the advantage of pairing, an increased P-value. However, there was a possible drawback of more elaborate designs: they would consume degrees of freedom so that, while the standard error being estimated was smaller, the estimate of it is less precise.

> A more precise comparison is obtainable by [pairing] only if the corresponding values of the two series are positively correlated, and only if they are correlated to a sufficient extent to counterbalance the loss of precision due to basing our estimate of variance upon fewer degrees of freedom.

Example 21 illustrates how pairing may bring a loss of precision. There are four paired observations and with three degrees of freedom the value of P is between .01 and .02. If the observations are unpaired, there are six degrees of freedom and the value of t is so large that it is beyond the range of the table, showing that P is "extremely small".

These considerations are not applied to design when that topic is discussed in Chapter VIII of the book but the question is raised in relation to the Latin square in Eden & Fisher (1929, p. 204). There the matter rested until 1935.

9 "The Information Lost in the Estimation of Error"

"The information lost in the estimation of error", the last section of the information chapter (XI) of the *Experiments* (1935, pp. 247–50), addresses the issue of precision and the number of degrees of freedom: "There is . . . one further point in connection with experiments involving measurements, to which the theory of errors is applicable, which may be cleared up by the methods of this chapter."

Fisher (p. 248) explains the issue in the language of "absolute precision",

> When, as a result of an experiment, a value x has been assigned a sampling variance, s^2, validly and correctly estimated from n degrees of freedom, the position is not the same as if the variance were known with exactitude. . . . We need . . . in considering the absolute precision

of an experimental result, to take into account, not only the estimate s^2 derived from the data, but also the number of degrees of freedom upon which our estimate, s^2, was based.

In this formulation, the "amount of information supplied by an observed value, x" about μ in the case of *known* σ^2 (given a normal density for x) is $\frac{1}{\sigma^2}$.

"The methods of this chapter" (Section 7 above) cannot be applied mechanically for now there are two parameters and, besides, there is scope for using the fiducial argument, which Fisher thought was inapplicable in the case of discrete sample spaces. He had first used the fiducial argument in 1930—see Aldrich (2000)—and he described the fiducial distribution for μ, obtained from the t-distribution for $(x-\mu)/s$, twice in 1935—in the *Experiments* (chapter X) and in "The Fiducial Argument in Statistical Inference". For a brief guide on matters fiducial, see Stone (1983).

Fisher (1935, p. 248) calculates the information given by the experiment about μ using the t-distribution in place of the normal curve, arguing fiducially:

> the probability that the quantity t, defined by the relationship
>
> $$x - \mu = st,$$
>
> where x is the observed value and μ the hypothetical value of which it is an estimate, shall lie in any assigned range, dt, is given by the formula
>
> $$df = \frac{\frac{n-1}{2}!}{\frac{n-2}{2}!\sqrt{\pi n}} \cdot \frac{dt}{\left(1+\frac{t^2}{n}\right)^{\frac{1}{2}(n+1)}}$$
>
> or in terms of x and μ, by
>
> $$df = \frac{\frac{n-1}{2}!}{\frac{n-2}{2}!s\sqrt{\pi n}} \cdot \frac{dx}{\left(1+\frac{(x-\mu)^2}{ns^2}\right)^{\frac{1}{2}(n+1)}}.$$

This derivation of a "*df* ... *dx*" density for x from the "*df* ... *dt*" density was distinct from Fisher's other fiducial efforts, the familiar "*df* ... $d\mu$" distribution for μ, also from 1935, and the "fiducial prediction" from a normal sample in *Statistical Methods and Scientific Inference* (1956, chapter V, Section 4)). The "*df* ... *dx*" argument does not seem to have been discussed in the literature on the fiducial argument.

To finish, Fisher does what he had been doing since the 1920s (Section 2), he uses the density for x to calculate the information relating to μ. This turns out to be

$$\frac{n+1}{(n+3)s^2}.$$

He comments on the result

> It appears that the true precision of our estimate is somewhat lower than it would have been, had the variance been known with exactitude to be s^2. In the extreme case, when $n = 1$, and the estimate is based on only 1 degree of freedom, the precision is halved. And in general, the true precision is less than it might be thought, if the uncertainty of our estimate of the variance were ignored, by the fraction $2/(n+3)$. It may thus be worth while to sacrifice, to some small extent, the aim of diminishing the value of s^2, if this diminution carries with it any undue reduction in the number of degrees of freedom, available for the estimation of error.

There are two concerns here: how is precision *achieved* by an experiment to be measured and how is the choice between experiments to be made. The second is a continuation of the precision and degrees of freedom theme discussed in the last section.

In the seventh edition of the *Methods* Fisher (1938, pp. 274–275) incorporated the correction in his reworking of the canonical analysis of 1925a (see Section 6) to find the "amounts of information elicited by different methods":

> In this case, as in many others, the lower mean square is obtained at the expense of some reduction of the number of degrees of freedom on which the estimate of error is based. This makes the test of significance somewhat less stringent. If n is the number of degrees of freedom for error, the loss of information due to this cause is found... to be the fraction $2/(n+3)$, so that taking this factor into consideration we may summarise the results as follows:–

	D.F. for Error	Units of Information/Crude	Adjusted
Randomisation of 20 plots	15	28.18	25.05
Randomisation in 4 blocks	12	51.07	44.26
Eliminating order in block	11	80.07	68.03

> Even when allowance is thus made for the degrees of freedom absorbed, it is clear that in this case both the use of blocks, and that of order within the block, have been extraordinarily profitable.

I have presented the construction of this table in two stages (the first was described in Section 6) but Fisher's readers got both stages together. It is possible that Fisher only went fully informational when he had found a resolution of the degrees of freedom puzzle.

The analysis provoked two rounds of private discussion 20 years apart. M.S. Bartlett who cared about information—see Aldrich (2005)—wrote to Fisher about the analysis in the *Experiments*. For a few weeks around the New Year of 1936, they tried unsuccessfully to make themselves understood; see Bennett (1990, pp. 46–51). Fisher returned to the subject in the 1950s, or rather it returned to him. In 1955 Finney put to him a "little paradox" that "the average of the precision is greater than if σ^2 were known" (Bennett, 1990, pp. 95–103). Finney added a P.S. to one of his letters, "I see no reason why you should spend time in putting me right, but I thought the paradox might amuse you" (p. 101). Fisher did not rise to the challenge, nor did he add anything about what he hoped to achieve with the measure.

Fisher reacted differently when he read John Nelder's review of Federer's *Experimental Design* (1955). Nelder (1956, p. 341) criticised the book for not evaluating the work it had précised, instancing its presentation of Fisher's measure for "the relative efficiency of two designs with errors based on different numbers of degrees of freedom". Federer was at fault because the measure depends on the fiducial argument "not everywhere accepted (and of a type which the author nowhere else uses)" and because "other intuitively reasonable measures give rather different results." Fisher felt *he* was being smeared and complained to Nelder. At the same time he wanted to know what those "other" measures were. Nelder and Finney (also consulted) referred him to Cochran & Cox (1950); see Bennett (1990, pp. 257–259, 280–283). Cochran & Cox (p. 18) mention the "complicated tables" in Neyman *et al.* (1935) and then give an argument "which though logically faulty, leads to an approximation that is good enough for most practical purposes." They also report Fisher's argument (p. 28).

Nelder's reply included a general objection to the fiducial argument. Fisher did not expect to be answered back and he dismissed Nelder like a schoolboy, who had not paid enough attention to his teachers, or who had been taught badly (Bennett, 1990, p. 282; see also Senn, 2003). All this activity had a precipitate in the next editions of the *Methods* and the *Experiments*. In the *Methods* Fisher (1958, p. 267) made a last addition to the analysis of 1925 (Section 6):

> It should be stated that the adjustment ... is correct and exact, since more complicated and inexact adjustments have been proposed by later writers who evidently have not understood the problem.

Fisher wrote at greater length in the *Experiments* (1960, p. 244). He made a new point, that the omnibus character of the adjustment/allowance made it an advance on what Gosset could have done in 1908 and what he (Fisher) had done in 1925 (Section 8):

> Writers averse to the use of fiducial probability (Neyman, 1935; Walsh, 1949) have attempted solution of this problem by other paths. The results are immensely complex, and these writers seem to overlook the fact that for *given* levels of significance, *Student* has already solved the problem in 1908. What is needed in experimental design is knowledge of the allowance to be made for the limitation of the number of degrees of freedom, before the level or levels of interest are known. The allowance found in this Section have been in use now for 25 years without any serious alternative. It is not an approximation, or a provisional artifice, and has been unintelligible only to those who over a long period resisted the cogency of the fiducial argument.

This section has followed Fisher's attempt to ground in an information measure the relation between degrees of freedom, precision and the sensitivity of tests, which had first exercised him in 1925. The point was "cleared up" to his own, if not to anybody else's, satisfaction.

10 Decisions: Economics and Polemics

Finally, we turn to the economic element in Fisher's teaching, how he had taught us to conduct experimental and observational inquiries so as "to maximise the information obtained for a given expenditure". We also examine his criticisms of Wald's decision theory and note some similarities between his work and that of others on comparing experiments.

In 1926 John Russell, the Rothamsted director, complained that many experiments "do not give as much information as might have been hoped, considering the cost"; he then showed "how experiments can be arranged to give the maximum return for the work and money spent on them" (p. 989). Whether Fisher took this line from Russell or vice versa is unclear for the methods Russell was describing were inspired by Fisher. On that day, however, "information" appears to have been a technically innocent term.

In the *Experiments* Fisher (1935, p. 189) wrote earnestly—and more concretely than in 1930 (Section 7 above)—about the need for "information costings":

> it is important that the items of labour and skilled supervision chargeable to a particular method of experimentation, shall be fairly and carefully recorded and calculated. For any time and labour devoted to experimental work must be regarded as having been diverted from other work of scientific value, to which they might otherwise have been given. Even rough costings of this kind will usually show that the efficiency with which limited resources can be applied, is capable of relatively enormous increases by careful planning of the experimental programme,

> and there is nothing in the nature of scientific work which requires that the allocation of the resources to the ends aimed at should be in any degree rougher, or less scrupulous, than in the case of a commercial business.

Fisher probably thought that a detailed cost analysis, as a piece of accounting, would have been out of place in the *Experiments*. But he indicates what is involved when he evaluates a year of uniformity trials prior to the experiment proper being done:

> It may be noted . . . that with annual agricultural crops, knowledge of the yields of the experimental area in a previous year under uniform treatment has not been found sufficiently to increase the precision to warrant the adoption of such uniformity trials as a preliminary to projected experiments. Such a procedure necessarily doubles the experimental labour, and as it is not found to double the amount of information but to increase it, perhaps by 50 per cent., it is clearly unprofitable. For, by the application of twice the expenditure in time and money in the experimental year, the amount of information recovered may with confidence be expected to be approximately doubled (1935, p. 166).

The parallel discussion in the *Methods* appears in the section on the analysis of covariance, the technique by which the previous year's experience would be used. There Fisher (1934, p. 271) remarks, "The chief advantage of the analysis of covariance lies . . . in the guidance it is capable of giving in the design of an observational programme, and in the choice of which of many concomitant observations shall in fact be recorded."

The economic theme, unlike the information theme, was uncontroversial for it was regarded as a truism that design mattered because resources were limited—it was the premise for Gosset's work, described in Section 4 above. Fisher did not develop a rich *economics* of experimentation because he believed that, for *given* expenditure, there was a best way of increasing information; all that was needed was a calculation giving the information yield of each design. There would have been more to say if he had asked how much the information is *worth* but, as we now see, he held such an evaluation impossible in the case of scientific research.

Fisher reacted strongly to the statistical decision theory of Wald. He attacked *Statistical Decision Functions* (1950) in an article, "Statistical Methods and Scientific Induction" (1955), in a book *Statistical Methods and Scientific Inference* (1956) and in passages he added to the *Methods* (1958) and the *Experiments* (1960). Only in the article (1955, p. 70) did he mention Wald's version of the spanning formula, "the treatment of the design of experimentation as a part of the general decision problem", but he was too outraged at the idea of decision theory and at Wald's ignorance in writing about design without mentioning replication, control or randomization (the three "elements") to discuss Waldian spanning.

Fisher's criticisms sometimes have a surreal quality, as when he (1956, p. 100) conjures a dystopia where "a giant computer programmed with Decision Functions" replaces "responsible and independent" scientific thinkers and the Cold War is not far away when he writes about the totalitarian potential of decision theory (see Marks, 2003). Equally remarkable is the standpoint of the purest of pure scientist. Although in 1956 Fisher's stint at Rothamsted was more than 20 years in the past, he continued to follow the work there and to identify with its objectives. The point of testing varieties and fertilizers was to find better yielding varieties and more effective fertilizers. "The object . . . is to find out which will pay the farmer best" declared 'Student' in his "On Testing Varieties of Cereals" (1923, p. 271). Treating expenditure on experiments as an investment, Gosset estimated that new varieties of barley in Ireland had produced a gain of not less £250,000 per year and "as the cost of experiments from the commencement to the present time cannot have reached £40,000 the money has been well spent." Gosset was in business but

Fisher (1926, p. 503) was not above considering the confidence of the "purchasing public" when establishing the effectiveness of a manure.

Statistical Methods and Scientific Inference has the familiar phrases about experiments, information and resource cost (pp. 5, 139) but design is not treated in detail. Decision theory is treated in the chapter, "Some Misapprehensions about Tests of Significance." Fisher (1956, pp. 75–78) begins by arguing that the Neyman–Wald account of testing is based on a false analogy between "tests of significance" and "acceptance decisions". Tests of significance were originally means by which the "the research worker gains a better understanding of his experimental material." However the tests have been reinterpreted "as means to making decisions in an acceptance procedure." In Fisher's estimation the differences between the two situations are "many and wide". One of the "deepest" lies in the population: for the quality controller, the population has an unequivocal existence as a particular batch of goods, to be accepted or rejected as substandard, while in scientific inference the population is the "product of the statistician's imagination" (p. 77). Most of the chapter—and all of the analysis—is given to discussing what that population should be in the case of regression, the 2×2 table and the Behrens–Fisher problem. See Aldrich (2005) for this theme.

The other "dissimilarity" is in the use of a loss function for making the wrong decision. "It is important that the scientific worker introduces no cost functions for faulty decisions, as it is reasonable and often necessary to do with an Acceptance Procedure. To do so would imply that the purposes to which new knowledge was to be put were known and capable of evaluation" (p. 102, cf. 1955, p. 77). Gauss, who Fisher greatly admired and whose goals were as scientific, used a loss function in his second theory of estimation.

An interesting sidelight on Fisher's decision theory is provided by Lindley's "On a Measure of Information Provided by an Experiment" (1956). This paper belongs to the comparison of experiments literature which Blackwell (1951) and others developed from Wald's theory; for a recent review see Goel & Ginebra (2003). Lindley combined the Shannon (1948) notion of information and Bayesian decision theory by putting information into the utility function. Lindley (p. 999) refers to Fisher, the critic of decision theory, but not to the theorist of information and experiments. With a nod to Fisher (1955), Lindley suggests the object in experimentation is "not to reach decisions but rather to gain knowledge about the world." Replacing the decision objective by an inference objective, Lindley measures success by expected information gain. Lindley's (1956, p. 987) maxim "perform that experiment for which the expected gain in information is the greatest and continue experimentation until a preassigned amount of information has been attained" sounds like a close relative of Fisher's "to conduct experimental and observational inquiries so as to maximise the information obtained for a given expenditure." While Lindley mentions Fisher's objections to the view that "modern statistics *is* decision theory" and to the interpretation of significance tests as Wald decision problems, he does not mention that Fisher had views on the "measure of information provided by an experiment." "Fisher information" was adopted as a technical tool, for example, by Kiefer (1959), but with no awareness that Fisher had used the tool to compare experiments.

11 Summary and Conclusions

This investigation started from a description of where statistics stood in the mid-20th century and a story of how it had got there. I have argued that description and story reflect an information turn in Fisher's thought on experiments in the course of the 1930s. I now summarize the argument and speculate on why the development left so little mark on the literature. The only paper I know linking Fisher's ideas on design and his ideas on estimation, Seidenfeld (1992), does not mention that Fisher made his own links.

One reason why Fisher's information turn had so little effect was that statisticians were turning in other directions. They were following Neyman (1935, p. 75) in his judgement that the "frequency of errors in judgement" provides a "sufficiently simple and unquestionable principle in statistical work" while the "amount of information is too complicated and remote to serve as a principle". Fisher's information addition to Gosset's (and his own) standard deviation was the proverbial fifth wheel. In this connection it is worth noting that when Finney (1955, p. 16) treated Fisher's situations of a "more general character" (those described in Section 7 above) he worked entirely in terms of standard errors.

Statisticians were not interested in looking but was there much to see? There was some stamping of INFORMATION on old work and a flow of aphorisms but only two new technical contributions described in Section 7 and 9. These were noticed: one had a life in the specialized world of genetics—see Mather (1938)—and the other went into textbooks on design. The *Methods* was an old book and the contributions were additions to existing analyses but the *Experiments* was new and could have been built around information. However, there were other things the author wanted to emphasize. The information principle never found expression in a structure comparable to Wald's *Statistical Decision Functions* or Raiffa and Schlaifer's *Applied Statistical Decision Theory.* Fisher's last book, *Statistical Methods and Scientific Inference* (1956), emphasized the *variety* of inferences and information is confined to the chapter on the principles of estimation

To say that Fisher faced an unreceptive audience, or that he did not write a book on information, is not to say that the statements, or the work behind them, had no significance for him. In the 1920s he had been hugely impressed by what he had discovered about information; his "reduction of data" definition of statistics (Section 2) applied his deepest insight in the broadest possible way. The spanning formula of 1951 was a natural development of this definition to accommodate the design of experiments. The two technical contributions were a design addition to his canonical statement on estimation and an information addition to his canonical statement on the advantage of design. Fisher's practice was to use examples to make general points and these were important additions to major examples. The work described in Section 9 attempting to link information and testing had significance for Fisher, even if his readers were unpersuaded by it.

The story and the spanning formula date from 1947/51 but by 1935 they would have seemed to Fisher natural ways of expressing his achievement. His account of how "armed with amount of information as a practical tool, statisticians came to study what forms of experiment . . . would yield the most information" is literally false, given that so much of his work on experiments was done without the apparent benefit of this "practical tool". Yet, when he started to bring information into experiments it was with the understanding that two of the "three elements" of replication, local control and randomization were manifestations of the information principle. He had been using the tool without giving it its proper name.

Acknowledgement

I am grateful to the editor and a referee for their suggestions.

References

Aldrich, J. (1997). R.A. Fisher and the making of maximum likelihood 1912–22. *Stat. Sci.*, **12**, 162–176.

Aldrich, J. (2000). Fisher's 'Inverse Probability' of 1930. *Int. Stat. Rev.*, **68**, 155–172.

Aldrich, J. (2005). Fisher and regression. *Stat. Sci.*, **20**, 401–417.

Atkinson, A.C. & Bailey, R.A. (2001). One hundred years of the design of experiments on and off the pages of "Biometrika". *Biometrika*, **88**, 53–97.

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

Blackwell, D. (1951). Comparison of experiments. *Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 93–102. Berkeley: University of California Press.

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

Box, J.F. (1980). R.A. Fisher and the design of experiments, 1922–26. *Amer. Stat.*, **34**, 1–7.

Cochran, W.G. & Cox, G.M. (1950). *Experimental Designs*. New York: Wiley.

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

Engledow, F.L. (1925). Theory of experimental error. *J. Min. Agric.*, **32**, 326–333.

Federer, W.T. (1955). *Experimental Design: Theory and Application*. New York: Macmillan.

Fienberg, S.E. & Hinkley, D.V. (1980). *R.A. Fisher: An Appreciation*. New York: Springer.

Finney, D.J. (1960). An introduction to the theory of experimental design. Chicago: Chicago University Press.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc., A*, **222**, 309–368.

Fisher, R.A. (1924). On a distribution yielding the error functions of several well known statistics. *Proc. International Congress of Mathematics, Toronto*, **2**, 805–813.

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

Fisher, R.A. (1925b). Theory of statistical estimation. *Proc. Cambridge Philosophical Society*, **22**, 700–725.

Fisher, R.A. (1926). The arrangement of field experiments. *J. Min. Agric.*, **33**, 503–513.

Fisher, R.A. (1929). Statistics in biological research, Letter in *Nature*, August 17, 266–267. Reproduced with notes in Aldrich (2004).

Fisher, R.A. (1930). *Statistical Methods for Research Workers*, 3rd edition. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1932). *Statistical Methods for Research Workers*, 4th edition. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1933). The contributions of Rothamsted to the development of the science of statistics. *Annual Report of the Rothamsted Experimental Station*, pp. 43–50.

Fisher, R.A. (1934). Two new properties of mathematical likelihood. *Proc. Royal Society, A*, **144**, 285–307.

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

Fisher, R.A. (1938). *Statistical Methods for Research Workers*, 7th edition. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1947). Development of the theory of experimental design. *Proc. International Statistical Conferences* (Washington), **3**, 434–439.

Fisher, R.A. (1951). Statistics. In *Scientific Thought in the Twentieth Century*, Ed. A.E. Heath. London: Watts.

Fisher, R.A. (1955). Statistical methods and scientific induction. *J. R. Stat. Soc., B*, **17**, 69–78.

Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1958). *Statistical Methods for Research Workers*, 13th edition. Edinburgh: Oliver & Boyd.

Fisher, R.A. (1960). *The Design of Experiments*, 7th edition. Edinburgh: Oliver & Boyd.

Fisher, R.A. & Eden, T. (1929). Studies in crop variation. VI. Experiments on the response of the potato to potash and nitrogen. *J. Agric. Sci.*, **19**, 201–213.

Fisher, R.A. & Wishart, J. (1930). The arrangement of field experiments and the statistical reduction of the results. *Imperial Bureau of Soil Science, Technical Communication*, **10**.

Goel, P.K. & Ginebra, J. (2003). When is one experiment 'always better than' another. *Statistician*, **52**, 515–537.

Hinkley, D.V. (1980). Theory of statistical estimation: The 1925 Paper. In *R.A. Fisher: An Appreciation*. Eds. S.E. Fienberg and D.V. Hinkley, (1980) pp. 685–694. New York: Springer.

Hinkley, D.V. (1980a). Fisher's Development of Conditional Inference. In *R.A. Fisher: An Appreciation*. Eds. S.E. Fienberg and D.V. Hinkley (1980) pp. 6101–108. New York: Springer.

Hinkley, D.V. (1980b). Contribution to the discussion of Basu's "Randomization Analysis of Experimental Data: The Fisher Randomization Test". *J. Amer. Stat. Ass.*, **75**, 582–584.

Kiefer, J. (1959). Optimum Experimental Designs (with discussion). *J. R. Stat. Soc., B*, **21**, 272–319.

Kempthorne, O. (1952). *The Design and Analysis of Experiments*. New York: Wiley.

Kempthorne, O. (1966). Some Aspects of Experimental Inference. *J. Amer. Stat. Ass.*, **61**, 11–34.

Lindley, D.V. (1956). On a measure of information provided by an experiment. *Ann. Math. Stat.*, **27**, 986–1005.

McMullen, L. (Ed.) (1970). *Letters from W.S. Gosset to R.A. Fisher 1915–1936* with summaries by R.A. Fisher with a foreword by L. McMullen, printed by Arthur Guinness for private circulation.
Marks, H.M. (2003). Rigorous uncertainty: why R.A. Fisher is important. *Int. J. Epidemiol.*, **32**, 932–937.
Mather, K. (1938). *The Measurement of Linkage in Heredity*. London: Methuen.
Mercer, W.B. & Hall, A.D. (1911). The experimental error of field trials. *J. Agric. Sci.*, **4**, 107–132.
Nelder, J.A. (1956). Review of *Experimental Design, Theory and Application* by Walter T. Federer. *J. R. Stat. Soc., A*, **119**, 340–341.
Neyman, J. (1935). Discussion of Fisher (1935a). *J. R. Stat. Soc.*, **98**, 73–76.
Neyman, J., Iwaszkiewicz, K. & Kołodziejczyk, St. (1935). Statistical problems in agricultural experimentation. *J. R. Stat. Soc. Supplement*, **2**, 107–154.
Pearce, S.C. (1979). Experimental design: R.A. Fisher and some modern rivals. *Statistician*, **28**, 153–161.
Pearson, E.S. (1990). '*Student', A Statistical Biography of William Sealy Gosset*, Edited and Augmented by R.L. Plackett with the Assistance of G.A. Barnard. Oxford: Oxford University Press.
Preece, D.A. (1990). R.A. Fisher and experimental design: A review. *Biometrics*, **46**, 925–935.
Raiffa, H.A. & Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Boston: Graduate School of Business Administration, Harvard University.
Russell, J. (1926). Field experiments: how they are made and what they are for. *J. Min. Agric.*, **32**, 989–1001.
Seidenfeld, T. (1992). R.A. Fisher on the Design of Experiments and Statistical Estimation. In *The Founders of Evolutionary Genetics*, Ed. S. Sarkar. Dordrecht: Kluwer.
Senn, S. (2003). A Conversation with John Nelder. *Stat. Sci.*, **18**, 118–131.
Shannon, C.E. (1948). A mathematical theory of communication. *Bell Sys. Tech. J.*, **27**, 379–423 & 623–656.
Stone, M. (1983). Fiducial Probability. In *Encyclopedia of Statistical Science volume 3*, Eds. S. Kotz and N.L. Johnson pp. 81–85. New York: Wiley.
Street, D.J. (1990). Fisher's contributions to agricultural statistics. *Biometrics*, **46**, 937–945.
'Student' (1908). The probable error of a mean. *Biometrika*, **6**, 1–25.
'Student' (1911). Note on a method of arranging plots so as to utilize a given area of land to the best advantage in testing two varieties, Appendix to Mercer and Hall's "The Experimental Error of Field Trials". *J. Agric. Sci.*, **4**, 128.
'Student' (1923). On testing varieties of cereals. *Biometrika*, **15**, 271–293. Amendment and correction, **16**, 1924, p. 411.
Wald, A. (1950). *Statistical Decision Functions*. New York: Wiley.
Walsh, J.E. (1949). On the "Information" lost by using a *t*-test when the population variance is known. *J. Amer. Stat. Assoc.*, **44**, 122–125.
Welch, B.L. (1937). On the *z*-test in randomized blocks and Latin squares. *Biometrika*, **29**, 21–52.
Yates, F. (1964). Sir Ronald Fisher and the design of experiments. *Biometrics*, **20**, 307–321.
Youden, W.J. (1951). The Fisherian revolution in methods of experimentation. *J. Amer. Stat. Assoc.*, **46**, 47–50.

Résumé

En 1951 R.A. Fisher décrivait ce qui avait été accompli jusque là au 20ème siécle: "nous avons appris (i) à conserver dans sa réduction statistique l'information scientifique latente dans tout ensemble d'observations (ii) de conduire des investigations expérimentales et observationnelles de façon à maximiser l'information obtenue pour une dépense donnée." Cet article se demande ce que Fisher voulait dire et, en particulier, comment il voyait son travail de conception expérimentale comme contribution à l'objectif de maximiser l'information pour une dépense donnée. Les matériaux examinés vont des travaux détaillés sur des questions telles que "l'information perdue dans la mesure de l'erreur" jusqu'aux polémiques contre la théorie de la décision.

[*Received March 2005, accepted May 2007*]