R.A. Fisher and the Design of Experiments, 1922–1926

JOAN FISHER BOX*

This article traces the development of the design of experiments from origins in the mind and professional experience of R.A. Fisher between 1922 and 1926. The article indicates how the analysis of variance procedure stimulated design, being justified by the principle of randomization that Fisher introduced with the analysis, and exploited by his use of blocking and replication. The article indicates the radically new form and efficiency of factorial block designs, shows the further advantages accruing to factorial arrangements through confounding, and suggests how Fisher's close collaboration with experimenters stimulated these developments.

KEY WORDS: Analysis of variance; Confounding; Design of experiments; Factorial block design; Randomization; Replication.

1. INTRODUCTION

In 1926 an article by Sir John Russell was published under the title, "Field Experiments: How They Are Made and What They Are." By the standards of the day Russell was well qualified to speak on the subject. An experienced agricultural chemist, he had been head of the chemistry department at Wye College before moving to Rothamsted Experimental Station in 1907. There he continued his research, and in 1912 was chosen to succeed A.D. Hall as director. Russell continued to keep close contact with the experimental work in the laboratories and on the farm at Rothamsted. Ideally, he envisaged "collecting a team of scientific workers and putting them on an experimental farm, where they could apply their various scientific disciplines" (Thornton 1966, p. 460). He was able to realize this idea when expansion of the work became possible after the war.

Russell's article on field experiments (1926) exhibits the state of the art of experimental design as it was generally understood in 1926. It provoked an immediate response from Russell's statistician at Rothamsted,

realize this idea when expansion of the work became possible after the war.

Russell's article on field experiments (1926) exhibits the state of the art of experimental design as it was generally understood in 1926. It provoked an immediate response from Russell's statistician at Rothamsted,

R.A. Fisher, whose article, "The Arrangement of Field Experiments," appeared in the same journal later that year. Apart from a few pages on design in *Statistical Methods for Research Workers* (1925), this was Fisher's first discussion of the design of experiments. In it the principles of randomization, replication, and blocking were enunciated, their application was exemplified using randomized Latin square and factorial block designs, and the possibilities of confounding were introduced. The aims of Fisher's experimental design were radically new: Whereas Russell's explicit aim was precision, Fisher's overriding considerations were validity and efficiency.

As a point of departure from which to follow Fisher's divergence from traditional views, one experiment is here considered in various aspects. This experiment was run at Rothamsted Farm in 1922 by the soil scientist Thomas Eden. It was analyzed by Fisher in "Studies of Crop Variation. II. The Manurial Response of Different Potato Varieties" (Fisher and Mackenzie 1923). In discussing this analysis with Eden, Fisher showed how profitably they could work together in planning experiments Eden was later to run. As a result of their collaboration, Eden gradually carried Fisher's ideas of design on to the field.

Fisher had been appointed statistician at Rothamsted Experimental Station in 1919. There had not previously been a statistician at Rothamsted, and Russell had only enough funds to pay his salary for six months, so the initial appointment was temporary. Russell hoped that six months would suffice to show whether a man trained in mathematics could be sufficiently useful in the analysis of Rothamsted data to warrant the creation of a permanent post for a statistician. As he explained later,

> On taking charge at Rothamsted I found great files of records which I knew I could never deal with adequately. . . . I knew that the Census authorities had methods for extracting information from great masses of data, and in 1919 after the war I applied both to Oxford and to Cambridge Universities for a young mathematician familiar with similar methods who would be prepared to examine our data and elicit further information that we had missed. (1966, p. 325)

This was the scope of the statistical duties envisaged for Fisher when he accepted the job. Fisher's expectations were similar. In 1947, he recalled,

> At the time, now about 35 years ago, when first I came to study statistical methods, nothing was further from my thoughts, or from those of my contemporaries, than that the art of experimental design would ever come to be, as it now surely is, an integral part of the subject. The horizon was indeed filled with problems of *fitting*, or as we should now say, of estimation. (p. 434)

> At the time, now about 35 years ago, when first I came to study statistical methods, nothing was further from my thoughts, or from those of my contemporaries, than that the art of experimental design would ever come to be, as it now surely is, an integral part of the subject. The horizon was indeed filled with problems of *fitting*, or as we should now say, of estimation. (p. 434)

His perception of his task changed when, in the course of his work on estimation, it was rigorously demonstrated that the amount of information extracted in the process of estimation could never exceed the quantity supplied by the data. As he explained,

> This finding, 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 statistician was no longer responsible for the accuracy or precision of the results of his labours. His business became much less like that of a conjuror who is expected to work wonders, and more like that of a chemist who undertakes to assay the proportion of gold in a sample. (1947, p. 435)

Thus by about 1922, Fisher recognized that a statistician, so long as he did his arithmetic right, had no responsibility for the value or the worthlessness of his estimates; consequently, "The weight of his responsibility was thrown back on to the processes by which the data had come into existence" (1947, p. 435). Fisher accepted the design of experiments as his charge.

* Joan Fisher Box, residing at 3437 Edgehill Pkwy., Madison, WI, is author of the scientific biography of her father, *R.A. Fisher, The Life of a Scientist* (1978).

Meanwhile he had been conjuring with the data, analyzing experiments, and, in the process, learning lessons that were to be important for design. His first task was to analyze crop records from one of Rothamsted's long-term experiments. In Broadbalk wheat field 13 plots (strips across the field) had been continuously under uniform treatment since 1852; the records of each plot for 67 consecutive years were available for comparison at the time Fisher did the analysis of the yield of dressed grain. Fisher observed slow changes in annual yield and his analysis was largely concerned with making allowance for them. The slow trends affected all the plots similarly, but they did not occur on other fields and could not be attributed to weather. To remove these trends, Fisher adopted and fitted orthogonal polynomials of fifth degree to the curves. For each plot, he was then able to make an analysis of variance due, as independent contributions, to progressive deterioration (with 1 degree of freedom), to slow trends (with 4 df), or to annual fluctuations (error, with $67 - 6 = 61$ df); and to test the significance of the deterioration by comparison with the residual error. This remarkable analysis is discussed more fully in Box (1978).

It should be added that in Fisher's account, the analysis of variance procedure is almost completely obscured. The tabulation (as shown here for one plot treated with farmyard manure) appears in an unfamiliar form, and the numerical quantities, which are in fact sums of squares, not mean squares, are not identified. Degrees of freedom are not tabulated, nor are they mentioned in the text; they are implied, however, because the significance level associated with probability (Pr) for slow changes is based on the appropriate number of degrees of freedom.

Plot	Annual Causes	Slow Changes	Deteri-oration	Total	Pr for Slow Changes
2b	33.2	17.6	.4	51.1	.000,002

2. RANDOMIZATION

The importance of the analysis of variance procedure in the development of Fisher's ideas of experimental design becomes apparent from "Studies in Crop Variation. II." (Fisher and Mackenzie 1923), in which the analysis of variance was made explicit, and the analysis of variance table appeared for the first time. In introducing this method of analysis, Fisher made it conditional on randomization. The experiment was essentially run in triplicate. Having divided the sums of squares of all the deviations from the general mean into two parts, one measuring the variation between parallel plots within triplicates similarly treated and the other the variation between means of the triplicates differently treated, he wrote, "If all the plots are undifferentiated, as if the numbers had been mixed up and written down in random order" (p. 315), then the average value of each of the two sums of squares would be proportional to the number of their respective degrees of freedom.

The statement rested on Fisher's understanding of the underlying distribution theory. His justification could have relied directly on normal theory assumptions if he had been prepared to assume that the observations had been drawn independently from a normally distributed population. But the assumption of independence was obviously not justified in any ordinary field experiment; observations of the fertility of adjacent plots were known to be not independent but highly correlated. Fisher perceived, however, that the random allocation of plot treatments would simulate the effect of independence in the distribution of the variance ratio, so that the analysis of variance and test of significance appropriate under normal theory assumptions would be approximately valid, *provided* the allocation of treatments to plots had been made deliberately at random.

Fisher tested this result informally, using data from uniformity trials. His confidence in the result, however, depended on the geometric representation that was by then second nature to him. He could picture the distribution of n results as a pattern in n-dimensional space, and he could see that randomization would produce a symmetry in that pattern rather like that produced by a kaleidoscope, and which approximated the required spherical symmetry available, in particular, from standard normal theory assumptions.

Thus, Fisher's first principle of experimental design arose at least partly from considerations not readily accessible to experimental scientists. The principle tied together what was done in the field and what could be learned from analysis of the results in a single logical package. What had been an empirical art of the experimenter was thus brought into the domain of the statistician; the role of the statistician was necessarily extended from the analysis of data to embrace the whole conduct of experimental inquiry.

Randomization was not readily accepted either by the mathematicians or by the experimenters (Box 1978, pp. 147–150; 262–270). Russell referred to it in 1926 as "a further refinement now being introduced at Rothamsted" (p. 1000); he then gave a correct argument for randomization as the guarantee of the validity of the estimate of error. But he continued: "In practice this [randomization] is impossible . . .; a compromise has to be made between what is desirable and what is practicable. The best practicable arrangement is to have as many repetitions as there are treatments, to set the plots out in chessboard fashion . . . [in] a 'Latin Square'." From Russell's earlier discussion and examples, it is clear that the Latin square was acceptable to him because, randomization or no, it retained the principle of balance in the design. His aim was to make the experiment as precise as possible. Therefore he advocated systematically balanced designs that could counteract fertility trends of the field. His simplest example was a row crop with two treatments applied to strips of the field in a sandwich arrangement, ABBAABBA. . . . With more numerous treatments of a row crop, each replication was differently ordered, but in such a way as to keep the average distance between strips that had been similarly treated the same for all treatments. The same principle was extended to a two-way arrangement of plots in a chessboard design, so that similar treatments should never be put closer to each other than necessary. With such designs the real error was usually reduced.

Arguing for randomization, Fisher (1926) pointed out that if the systematic arrangement resulted in smaller real errors, it must also result in an inflated estimate of error; if it resulted in larger real errors, the estimate of error would be correspondingly diminished. In either case the false estimate of error would be liable to vitiate the conclusions drawn from the experiment. As Fisher put it in correspondence, the experimenter games with the devil; he must be prepared by his layout to accommodate whatever pattern of soil fertilities the devil may have chosen in advance. A systematic arrangement is prepared to deal only with a certain sort of devilish plan. But the devil may have chosen any plan, even the one for which the systematic arrangement is least appropriate. To play this game with the greatest chance of success, the experimenter cannot afford to exclude the possibility of *any* possible arrangement of soil fertilities, and his best strategy is to equalize the chance that any treatment shall fall on any plot by determining it by chance himself. Then if all the plots with a particular treatment have higher yields, it may still be due to the devil's arrangement, but then and only then will the experimenter know how often his chance arrangement will coincide with the devil's.

Design of Experiments (1935, §5) opens with an example:

> A lady declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested.

The example was actually taken from Fisher's experience at teatime one day at Rothamsted in 1921 or 1922, when he drew a cup of tea from the urn, then added the milk and offered the cup to Miss B. Muriel Bristol. She refused it, maintaining that it made a difference if the milk was added first. At William Roach's suggestion, they proceeded at once to test her assertion. Roach prepared the tea cups, and recalls with pride the overwhelming success of the lady (who became his wife soon after), but he does not mention randomization. In contrast, Fisher's (1935, §9, 10) discussion emphasized, among the first considerations in this first example of design, that randomization, "the physical basis of the validity of the test," was "the essential safeguard" contained in the experimental procedure for the validity of the estimate of error, and thus for the test of significance by which the result of the experiment was to be judged. Moreover, he included similar discussions of randomization in relation to each class of design introduced later in the book (§20, 26, 34).

In 1926 it remained to be seen whether a randomized design could be as precise as a systematic one. That year Eden was running the first randomized block design; Fisher used it as an example even before the results were known. On analysis (Eden and Fisher 1927), this experiment proved admirably precise. Despite this success, many of the older experimenters continued to doubt the accuracy of randomized designs. As late as 1936, W.S. Gosset ("Student" 1936, p. 118) expressed the view that "Since the tendency of deliberate randomization is to increase the error, a balanced arrangement like the half-drill [a systematic sandwich design] is best."

3. REPLICATION AND BLOCKING

The need for replication was widely acknowledged before Fisher clarified its essential role. As Russell (1926, p. 995) put it, "Variations in soil can be overcome only by repeating the experiment on the same field at the same time. This is now well recognized, and duplicate experiments have long been the rule." Russell's examples of strip and chessboard designs actually contain triplicates of each treatment. Three or four replications were usual in contemporary designs.

Eden's design (Fisher and Mackenzie 1923) was conceived as triple replication, but was not wholly so. The aim was to inquire into the response of 12 different potato varieties to either of two potash fertilizers, with or without farmyard manure. The field was first divided into two equal areas, one of which received farmyard manure. Each half was then divided into 36 plots on which the 12 varieties were planted in triplicate in a chessboard arrangement. Finally, the plots were subdivided into three patches, which received either the basal dressing only or the basal dressing with either sulfate or chloride of potash. Some of the disadvantages of the design were obvious. When Gosset saw the analysis, he wrote to Fisher, "The experiment seems to me to be quite badly planned, you should give them a hand in that." In taking Gosset's advice, Fisher showed how many lessons in design such an experiment had to teach.

In his initial analysis of variance, Fisher made the mistake of using a single estimate of error for all the comparisons. He quickly discovered where he had gone wrong and published the correct analysis in *Statistical Methods for Research Workers* (1925, §42). Carefully explaining his reasoning, he derived the separate estimates of error and made the separate analyses required for the plots and for the patches. He considered data from one half of the field only, because without replication of the treatment on half-fields, no applicable estimate of error could be derived by which to assess the effect of farmyard manure.

Thereafter (1935, §26), he said of replication that "its main purpose, which there is no alternative method of achieving, is to supply an estimate of error."

Eden's experiment was not run as a randomized block design. None of Russell's (1926) examples was of this kind. Russell valued designs like the Knut-Vik systematic square and Fisher's randomized Latin squares because of the supposed increased precision obtained by balance. The advantages of block designs were not apparent until, with the coming of the analysis of variance, it became possible to isolate the variance to be ascribed to block differences and see how it could be eliminated in the analysis (Fisher 1935, §33). With Fisher's randomized Latin squares the variance could be eliminated in two directions at once, but these designs were restricted by the requirement that the number of replications had to equal the number of treatments compared. Randomized block designs, however, were not similarly restricted. Each block could be a small and compact arrangement on relatively homogeneous land, and the number of replications on different blocks could be increased to any desired amount without adversely affecting the precision of the experimental results. Fisher's (1926) example used eightfold replication in blocks of 12 treatments each. In *Design of Experiments* (1935, §25) Fisher explained the ways in which a larger number of replications in blocks served to diminish the error.

4. LATIN SQUARES

The Latin squares that Fisher used as experimental designs had been studied by mathematicians because of their importance in combinatorics. If square designs were to be properly randomized, it was necessary to enumerate the kinds of squares. In 1924 Fisher wrote to the authority on combinatorial analysis, P.A. MacMahon, to inquire about the method by which he had enumerated the 4×4 and 5×5 squares. Fisher wrote:

> It is possible that your method of solution will solve for me at least the first of the outstanding questions:
> (1) What experimental techniques of filling up the square will give each solution an equal chance of appearing?
> (2) Is such a technique necessary or sufficient for the statistical validity of an agricultural experiment?

Evidently, he found for himself satisfactory answers to both his questions. Within a month, he evolved his own direct method of enumeration and found for the 5×5 squares 56 symmetrical pairs of reduced squares, of which MacMahon had missed 4. A few months later he enumerated the 6×6 squares, incidentally demonstrating the truth of Euler's belief that there are no 6×6 Greco-Latin squares. Having made the enumerations, he proceeded at once to introduce the squares as experimental layouts. Later, on checking Fisher's enumeration of the 6×6 squares, F. Yates found some squares Fisher had missed; the enumeration was corrected and published (Fisher and Yates 1934).

Fisher found the work of enumeration fascinating. It appealed to his curiosity and to his mathematical ingenuity. He was delighted too by the individual character some of the sets displayed. In addition, there was to be considerable experimental spin-off from the purely theoretical work, for the combinatorial properties of the square arrangements soon came into play in the design of factorial arrangements, the exploitation of Greco-Latin squares of various sizes, and the incomplete block designs that F. Yates invented. Fisher had fun with enumeration and, as always with his pleasures, he wanted others to share his enjoyment, so he used combinatorics to create unique building blocks for his children. From the wood yard across the back fence he obtained 24 wooden bricks, 6 in. $\times$ 3 in. $\times$ 2 in. in size, with distinct faces, edges, and ends. These he painted, a red surface being always opposite yellow, black opposite white, and blue opposite green, and no two bricks alike. When he saw how much the two little boys enjoyed their new toys, he doubled the number of bricks; later, when younger children joined the play, he redoubled the number to realize the 96 possible arrangements of colors on the bricks. By the time his youngest child was born in 1935, the oldest was rising 16 and able to organize the repainting of the worn surfaces and to appreciate the logical aspects of the process that obviously still gave his father immense pleasure.

The first of Fisher's Latin squares was laid out in 1924 at the Bagshot Nursery of the Forestry Commission. In 1926 Fisher offered to supply random Latin square designs to experimenters on request and to analyze the results of the experiments at Rothamsted. "For the purpose of variety trials and of those simple types of manurial trial in which every possible comparison is of equal importance," he found the Latin square "by far the most efficient arrangement" and the 5×5 and 6×6 squares "by far the most useful sizes" (1926, p. 510). In *Design of Experiments* (§33) he went so far as to assert:

> If experimentation were only concerned with the comparison of four to eight treatments or varieties the Latin Square would, therefore, be not merely the principal but almost the universal design employed. . . . Where it fails is to provide a means of testing simultaneously a large number of different treatments or varieties.

To obtain precision in such experiments a different type of arrangement was required and developed.

5. FACTORIAL BLOCK DESIGNS

In 1926 Russell wrote:

> The first essential is to frame a perfectly clear idea of what is expected from the experiment. An experiment is simply a question put to nature in the hope of discovering some secret. Even in the best-planned experiments, the answer can usually be only "yes" or "no," while, if the experiment is badly planned, no answer can be given. . . . The chief requirement is simplicity: only one question should be asked at a time. (p. 989)

In practice, however, although the experimenter might be interested in a single problem, he nevertheless might need to ask a series of dichotomous questions to solve the problem. If, for example, the problem concerned application of nitrogen as ammonium salts to oats, the salts could be applied in the form of either muriate or sulfate, in single or double amounts, and early or late in the season. Fisher argued that the experimenter should ask these questions simultaneously, because the response to one treatment might be conditioned by others. For example, the double dressing might be advantageous only if it were applied early. These were, in fact, the questions asked in Fisher's first factorial block design, which was being carried out by Eden on Rothamsted Farm at the time Fisher used it to illustrate the factorial method in his 1926 paper.

In answer to Russell's plea for simplicity, Fisher wrote:

> No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions or, ideally, one question at a time. The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed. (1926, p. 511)

The sort of questionnaire Fisher proposed was modeled on nature itself, the questionnaire posed by sets of often dichotomous genes, in all combinations, and answered by natural selection. Indeed, Fisher (1952, p. 3) traced the factorial idea to its origins in his genetic thought: "The factorial method of experimentation, now of lively concern so far afield as the psychologists or the industrial chemists, derives its structure, and its name, from the simultaneous inheritance of Mendelian factors."

This attribution illustrates how his wide interests as a statistician were fertilized by his equally wide interests as a geneticist concerned with human improvement. He had been still an undergraduate when his initiative resulted in the formation of the Cambridge University Eugenics Society. Already, in 1911, in addressing the undergraduate committee, he had assumed the multifactorial inheritance of continuously variable human traits. He had been led to the analysis of variance components (1918) as a means of demonstrating that multifactorial inheritance in accordance with Mendelian principles would result in precisely the sort of correlations found to exist between human relatives. Moreover, he had been much interested when genetic breeding had demonstrated that the effect of a particular gene could be greatly modified under human selection through its interaction with modifying genes at other loci. With such genic interaction clearly in his mind, Fisher found that the parallel between genetical and experimental situations was suggestively close. If two alleles (alternative genes at a particular locus) affected a continuous variable, like height, they would contribute to the outcome in just the same way that an experimental treatment at two levels might affect a continuous variable, like yield, of the field. With three pairs of alleles, or pairs of treatments, the effects might be simply additive, in which case the three pairs would make three independent contributions to the population variance. If there were interactions between the pairs, there would be, in addition, contributions from any or all of three two-factor interactions and possibly a further contribution from the interaction of all three. If the genes could have been identified and imposed like agricultural treatments, each of these effects and interactions could have been estimated by the analysis of variance.

Russell (1926) used a design he attributed to Fisher to exemplify an elaboration of a sandwich design by cross dressing. The design was a systematic arrangement of a 2^2 factorial run in duplicate. Eight strips were treated with phosphate as two sandwiches, thus: l p p l l p p l. The field was then divided across the strips, and nitrogenous fertilizer applied to opposite ends of the two sandwiches to produce the arrangement:

n	np	np	n	l	p	p	l
l	p	p	l	n	np	np	n

Russell was probably mistaken in attributing the design to Fisher. Given that the sandwich arrangement was to be run anyway, Fisher might have suggested the cross-dressing in order to provide the experimenters with desired comparisons, but having realized in 1923 that the validity of his estimates was conditional on randomization, he could not later have approved, let alone have suggested, such a systematic arrangement. In 1926 the design was in no way one he would have cared to have associated with his name.

Fisher's (1926) example of the factorial design consisted of eight randomized blocks, each containing the 12 treatments of a $3 \times 2 \times 2$ factorial. The three ammonia treatments were muriate application, sulfate application, and application of neither. These treatments were run with all combinations of early or late treatment in single or double quantity. The experimental layout was unusually large and complex but, as is fully set forth in the discussion of this experiment and its results (Eden and Fisher 1927), the disadvantages of the layout were outweighed by its efficiency.

1. All the comparisons were made between plots of the same size and precision and required only the one estimate of error between plots.
2. The validity of this estimate was guaranteed by randomization of the treatments on the plots of each block, the particular arrangement of each block having been determined separately by a chance process.
3. Eightfold replication greatly enhanced the accuracy of the experiment.
4. Each replication was contained within a separate block. Thus, every comparison was made between treatments within a single block, which contained only such soil heterogeneity as existed in the compact area

of 12 plots. The block size was chosen to suit the field conditions.

5. Because of the hidden replication within the full factorial, every comparison was made not between single plots but between the average of one set of four plots and the average of another set of four. By extension (Fisher 1935, §41), in some cases this hidden replication would be so abundant that overt replication might be dispensed with altogether.

6. Each of the three main effects could be estimated separately by comparison of the four plots that received the treatment with the four that did not receive it. The three two-factor interactions (indicating any difference in the effect of one treatment made by the presence of another) could be independently estimated by comparing the sets of four plots that differed in respect of the interaction term. Finally, the three-factor interaction could be estimated independently of all other effects and interactions by comparison of the appropriate sets of four plots. Every observation was used many times over in different combinations to estimate the seven main effects and interactions for each of the ammonium salts.

7. In addition to the immense efficiency of the factorial in simultaneously providing answers to many questions, most of which could not even have been asked using one question at a time, the factorial block arrangement had the added advantage of providing a wider inductive basis for conclusions drawn from it. The experimental treatments had all been applied under a variety of different conditions of the other experimental variables and of the several blocks. The experimenter could therefore have greater confidence in drawing general conclusions from a particular field trial. He might deliberately broaden this inductive basis by laying out blocks of a single experiment on different fields or even on different farms to compare the results on soils with different textures or different histories of cultivation, for example.

6. CONFOUNDING

Fisher's (1926) paper closed by introducing the idea of confounding. There was an inherent tug-of-war in the factorial block design: The larger the factorial, the more factors it contained, the more efficient it became (that is, a larger number of factors could be evaluated with the same precision by means of a smaller fraction of the number of observations that would otherwise be necessary); but the larger the block, the more soil heterogeneity it encountered. Fisher perceived this tug-of-war and also perceived that the factorial, by its very form, offered a way out of the difficulty.

The case he considered is a classic of manurial trials, experimental treatment with nitrogen, phosphate, and potash in a 2^3 factorial. With such an experiment, while the experimenter might be interested primarily in the three main effects, he might also wish to observe any two-factor interactions that occurred, but be fairly confident that the three-factor interaction was negligible. In such a case, Fisher suggested, he could run the experiment not in blocks of eight treatments but in smaller blocks of size four, each pair of smaller blocks contrasting those treatments normally contrasted in estimating the three-factor interaction. The three-factor interaction would thus be "confounded" with the difference between the two small blocks and eliminated as such in the analysis. All the other comparisons would be unaffected and remain unconfounded.

The possibility of discovering different confounding schemes to fit each new experimental situation was explored and exploited for many years after. It is still an active research field. It was possible to employ double confounding in cases in which two sorts of heterogeneity existed, as in the Latin square, with certain treatment effects confounded with rows and others with columns. The idea could be generalized to include more than two sorts of heterogeneity using Greco-Latin squares. Partial confounding used the idea of confounding different effects in different replications; thus effects confounded in one part of the experiment were unconfounded somewhere else. There was no end to the neat solutions possible. As Fisher (1935, §46) wrote enticingly, "The reader will find great advantages in investigating the possibilities. . . . The variety of the subject is, in fact, unlimited, and probably many valuable possibilities remain to be discovered." (In the third edition published in 1942, two new sections on confounding were included.) Fisher's skill in combinatorial manipulations and his pleasure in solving puzzles found new scope in contriving the beautiful and elaborate designs his colleagues needed to conserve all the advantages of the factorial arrangement while conforming to the block size appropriate to the field.

7. COLLABORATION IN THE FIELD

A report (Eden and Fisher 1929) on a series of manurial trials on potatoes tells its own story of the interplay between practical needs and theoretical solutions in the rapid development of ideas on the design of agricultural experiments.

In 1925 two experiments were run, a systematic 4×4 square to test qualitative differences, and a systematic arrangement of an incomplete factorial, in four blocks or strips running the length of the field, to test quantitative differences. In 1926 no change was made in the square design. In the block design, although there were still only four blocks, they were arranged to quarter the field; they contained a complete factorial of 16 treatments, and the treatments were randomized in each block. The results of this experiment, though improved, were rather inaccurate. More replications and smaller blocks were needed, but practical considerations forbade the use of a larger number of plots.

The problem was overcome by amalgamating the two experiments. In 1926 there had been 80 plots in all, 16 in the square and 64 in the four blocks of 16. In 1927 there were 81 plots, with nine blocks of 9 plots each. For this ninefold replication, the highest of the four levels of dressing both with nitrogen and with potash was dropped; there were now three levels of ammonium sulfate dressing to be tested at three levels of potash of three kinds: sulfate, muriate, and low-grade salt. Fisher treated the three levels of nitrogen and potash (the quantitative factors) as the basic block of nine treatments. He explained:

> The actual position of a plot considered only as representing potash and nitrogen interactions was determined entirely by chance. The element of chance also operated largely in the disposition of the qualitative factor, but there was one restriction. The restriction provided that any particular variety of potash manure should occur in the total of the nine blocks in conjunction with every amount of nitrogen three times. In every other way the distribution was at random. (1929, p. 210)

He pointed out that the amount of replication varied with each factor or interaction of factors concerned and listed the number of replications for each of the eight classes of comparison. He did not explain how he had arrived at this layout. The actual scheme of randomization was not important to his readers. It might merely have bewildered them to be told that it was in fact a 9×9 Latin square whose rows appear as the nine experimental blocks (treatment combinations at three levels of phosphate and three levels of nitrogen) and whose columns appear as the nine other factorial arrangements (of the three sorts of phosphate at three levels of nitrogen), which are necessarily orthogonal not only to each other but also to the blocks. Even today it comes as a surprise that so complex a design was planned the same year that Fisher's Latin square and factorial designs first saw print. What was important to his readers was to show how "The large and complex type of experiment finally adopted thus supplied more precise information on both heads [qualitative and quantitative] than could previously be obtained, and in addition to a more thorough exploration of the different combinations possible" (p. 213).

The size and complexity of the new designs was, at first, their most embarrassing feature. Russell (1926, p. 995) declared "no experiment should involve more than four or five" treatments. Of the 2^2 factorial in duplicate, he wrote: "The set involves 16 plots, but the agricultural operations can be managed without much difficulty" (p. 997); and of the Latin square: "Obviously the method requires a considerable number of plots. Its use at Rothamsted necessitates special arrangements for harvesting, thrashing, weighing and recording, which, however, are too intricate to be dealt with here" (p. 1000). Nevertheless, Russell gave Eden, and Fisher, a free hand, and allowed their complex designs to be run on Rothamsted Farm. Although doing so entailed many adjustments in field operations, the practical running of the experiments proved quite manageable, and in time the extra trouble came to seem a small price to pay for the highly satisfactory results.

[*Received October 1978. Revised June 1979.*]

REFERENCES

Box, Joan Fisher (1978), *R.A. Fisher, The Life of a Scientist*, New York: John Wiley & Sons.

Fisher, Ronald Aylmer (1971–1974), *Collected Papers of R.A. Fisher* (5 vols.), ed. J.H. Bennett, Adelaide: University of Adelaide. (See references to individual *Papers*).

——— (1951), *Contributions to Mathematical Statistics*, New York: John Wiley & Sons. (See references to individual *Contributions*.)

——— (1918), "The Correlation Between Relatives on the Supposition of Mendelian Inheritance," *Transactions of the Royal Society of Edinburgh*, 52, 399–433, and *Collected Papers* 1, no. 9.

——— (1921), "Studies in Crop Variation. I. An Examination of the Yield of Dressed Grain From Broadbalk," *Journal of Agricultural Science*, 11, 107–135, *Collected Papers* 1, no. 15, and *Contributions*, paper 3.

——— (1925), *Statistical Methods for Research Workers*, Edinburgh: Oliver and Boyd.

——— (1926), "The Arrangement of Field Experiments," *Journal of the Ministry of Agriculture of Great Britain*, 33, 503–513, *Collected Papers*, 2, no. 48, and *Contributions*, paper 17.

——— (1935), *Design of Experiments*, Edinburgh: Oliver and Boyd.

——— (1947), "The Development of the Theory of Experimental Design," *Proceedings of the International Statistical Conference*, 3, 434–439, and *Collected Papers*, 4, no. 212.

——— (1952), "Statistical Methods in Genetics," *Heredity*, 6, 1–12, and *Collected Papers*, 5, no. 248.

Fisher, Ronald Aylmer, and Eden, Thomas (1927), "Studies in Crop Variation. IV. The Experimental Determination of the Value of Top Dressings With Cereals," *Journal of Agricultural Science*, 17, 548–562, and *Collected Papers*, 2, no. 57.

——— (1929), "Studies in Crop Variation. VI. Experiments on the Response of the Potato to Potash and Nitrogen," *Journal of Agricultural Science*, 19, 201–213, *Collected Papers*, 2, no. 78, and *Contributions*, paper 18.

Fisher, Ronald Aylmer, and Mackenzie, W.A. (1923), "Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties," *Journal of Agricultural Science*, 13, 311–320, and *Collected Papers*, 1, no. 32.

Fisher, Ronald Aylmer, and Yates, Frank (1934), "The 6 × 6 Latin Squares," *Proceedings of the Cambridge Philosophical Society*, 30, 492–507, and *Collected Papers*, 3, no. 110.

Russell, Sir E. John (1926), "Field Experiments: How They Are Made and What They Are," *Journal of the Ministry of Agriculture of Great Britain*, 32, 989–1001.

——— (1966), *A History of Agricultural Science in Great Britain, 1620–1954*, London: Allen and Unwin.

"Student" (1936), "Cooperation in Large-Scale Experiments," *Journal of the Royal Statistical Society, Supplement*, 3, 115–122.

Thornton, H. Gerard (1966), "Edward John Russell 1872–1965," *Biographical Memoirs of Fellows of the Royal Society*, 12, 457–477.