Complex Experiments

By F. Yates, M.A.

[Read before the Industrial and Agricultural Research Section of the Royal Statistical Society, May 23rd, 1935, Sir William Dampier, Sc.D., F.R.S., in the Chair.]

1. *Introduction.*

Several papers have recently been read before this Section or published in the Supplement to the *Journal* which have dealt with one aspect or another of the randomized block and Latin square methods of carrying out replicated experiments.

These methods were first developed in connection with field trials in agriculture, but since their inception it has been abundantly clear that they are of very wide application, and are therefore of interest to experimental workers in almost all branches of science and technology. The paper I propose to give to-night may be considered as belonging to the same series. In it I intend to deal with another aspect of the methods which has not yet been discussed, namely the part which treatments play in experimental design.

Following the previous writers, I shall describe the special technique which has been developed in agricultural field trials, but it is hoped that the paper will prove of interest not only to agricultural workers, but also to workers in many other branches of research. For while the complex method has been most fully developed in connection with agriculture, the special difficulties occasioned by the necessity of eliminating soil heterogeneity both limit its application and complicate the method. It would appear that its use in other fields in which the material is more homogeneous would be even more fruitful.

In the absence of specialized knowledge of the problems and conditions which occur in other branches of research and industry it is impossible to give detailed examples of the utility of complex design in these fields. But a few general suggestions may be made to show the wide scope of the method. For the rest I would stress that there are many more ways than one of carrying out experimental work, and only with a knowledge of the basic principles of the available methods, and a good acquaintance with the sources of variation of the experimental material, can the experimenter hope to arrive at an efficient technique. For this reason it has appeared to me profitable to discuss somewhat fully the application of these methods to agricultural research. The difficulties to be overcome in other fields will not be identical, but many of them will be similar, and in so far as this is the case the solutions which have been arrived at in agriculture will be of material assistance. In so far as they differ fundamentally, new technique will be required, but such technique can only be evolved when in actual contact with the difficulties themselves.

As examples of particular applications of complex experimentation (or, as we shall call it, *factorial design*) to fields other than agriculture, I may first instance research into cotton-spinning problems carried out at the Shirley Institute. In some experiments described by Tippett in a paper recently read before this Section,[6] the adoption of complex methods enabled him to investigate three or four factors at once instead of one. His experimental problem was peculiarly simple, in that he was prepared to assume the virtual independence of the several factors he was investigating, so that the effect of variation in one factor could be regarded as substantially the same whatever the values of the other factors. (This method is described in Section 9.) In biological experiments such assumptions cannot be regarded as satisfactory, but I imagine that in many experiments on manufacturing processes they are more justified.

Similar problems might occur in almost any branch of manufacture. In testing different types of motor tyre, for instance, we might wish to vary both the speed and the load, or both and the surface; or in actual road tests types of car, drivers, average speeds and similar factors might be varied. Factorial design would result in considerably more efficient utilization of experimental resources.

Then there is the wide field of biological enquiry, of which the work I am about to describe is a special branch. I need only here instance dietetic experiments on animals, and especially on human beings (where experimental material is hard to come by). Here the interactions between the various components of any diet are of vital importance, and factorial designs would appear to be as necessary as they are in the very similar problems encountered in agronomic research.

As regards the applications of the designs alluded to in Section 10 (for which I shall propose the name of *incomplete randomized blocks*) I need only mention three instances. In human beings the resemblance between monozygotic twins is well known; if this is to be utilized experimentally we are limited to the equivalent of blocks of two. In certain virus experiments which involve the inoculation of leaves of young plants only five suitable leaves are growing on the plants at one time, so that blocks must be limited to five leaves each. And in many laboratory determinations in which variation in conditions seriously affects the results the number of determinations which can be carried out simultaneously (or approximately simultaneously) is strictly limited.

It has been thought better to retain throughout the terminology of agricultural field experiments, rather than create a more generalized terminology which might be applicable to all experimental material. This course recommends itself the more in that workers in any field will in practice refer to their experimental units by their appropriate names, so that some transposition of terms when passing from one field to another will always be necessary.

Since the whole evolution of experimental technique is bound up with questions of efficiency I have considered it worth while to devote a certain amount of space to a discussion of the actual gain in efficiency that is likely to result in agricultural field trials by the use of factorial design. This part of the paper will be of particular interest to agriculturists, since objections against factorial design have been made on the ground of actual loss of efficiency. But here again the discussion is likely to be of general interest, as indicating the lines to be followed in similar investigations in other fields.

I am afraid that certain parts of the paper will prove difficult reading to those not familiar with the principles of the analysis of variance. This is inevitable if undue bulk and tedious repetition are to be avoided. To those not acquainted with this branch of statistical technique I would recommend that they make a preliminary study of Fisher [1], and Fisher and Wishart [3], before they attempt to follow the analytical procedure of the examples given in this paper. I would also like to emphasize that a full appreciation of all the points of the various designs is more likely to be obtained if the actual examples given are worked over in full on a computing machine. It is to make this possible that I have reproduced here the numerical values of the individual plot yields.

Finally, it may be well here to emphasize two further points in connection with the evolution of an efficient experimental technique, which are sometimes lost sight of. The first is that the amount of work it is profitable to expend in developing a technique depends on the amount of experimental work of a given type that is likely to be undertaken. It is because the need for agricultural experimentation is so widespread and persistent that it has been worth while devoting considerable research merely to improving agricultural experimental technique. The other is that it is not necessarily any reflection on the ability of an experimenter if the methods he has employed and advocated are later found to be less efficient than other newer methods. In many fields of research there is a tendency for a sort of vested interest to grow up round an experimental method, which leads to its defence on entirely illogical and unreasonable grounds, and to a very grudging acceptance of newer methods.

2. *An Experimental Problem.*

Suppose that a new plant of agricultural importance is introduced into a country. (An example has recently been provided by the introduction of sugar beet into England.) What is the right way to set about determining the best varieties and the appropriate manurings and cultivations?

One procedure, extensively practised, is to divide the problem into a large number of smaller problems and attack each one separately. One set of experiments will be started to determine the best variety, a second set to determine the best manuring, a third the best cultivations. Nor need, nor does, the subdivision stop there. Responses to the three standard fertilizers, nitrogen, phosphate and potash, for instance, may be relegated to separate experiments.

This procedure has on the face of it a deceptive appearance of simplicity. The questions formulated are themselves simple:—Is one variety better than another? Is the yield increased by the application of nitrogen? Their answers can be obtained with an apparently high precision. But there is one very cogent objection. Clearly the experimenter on fertilizers, who, we will imagine, decides to confine his enquiries at the start to response to nitrogen, must choose some variety on which to experiment. He will probably choose what he considers is the best variety. After three years of experiment the experimenter on varieties may announce that some other variety is markedly superior. Are all the experiments on fertilizers now worthless, in that they apply to a variety that will no longer be grown? The experimenter on fertilizers will probably answer that the response of the two varieties is not likely to be widely different, and that his conclusions therefore still hold. But he has no experimental proof of this, only his experience of other crops, and would not have even this last if he and all other experimenters on fertilizers had persistently only experimented on one variety.

If the experimenter on varieties is so rash as to criticize his results on these grounds, however, and has himself laid down some standard of manuring for his varietal trials, the experimenter on fertilizers can effectively turn the tables by pointing out that the varietal trials are also of little value, being carried out at a level of manuring different from what he proposes to recommend.

Had the enquiries been combined into one system of experiments, so that all varieties were tested in conjunction with all levels of nitrogen, this imaginary controversy could not have arisen; for definite information would be obtained on whether all varieties did, in fact, respond equally to nitrogen. Moreover, if it was found that they did, a considerable gain in efficiency on the primary questions would result (provided that the experimental errors per plot were not greatly increased), since each plot would enter into both primary comparisons and would thus be used twice over. If, on the other hand, differential response was demonstrated, then although the response of the chosen variety would be known with less accuracy than if the whole experiment had been carried out with that variety, yet the experimenter might count himself lucky in that the possibility of false conclusions due to using another variety in his fertilizer trials had been avoided. Moreover, the conclusion as to the best variety might also require modification in the light of the differences in response to fertilizer.

When new varieties are being selected this greatly understates the advantage, for the essential and valuable difference of one variety over another may lie just in its ability to respond to heavy dressings of fertilizer; at the customary levels of manuring, it may be, the yields are about the same. Nor need this response be direct. In the case of wheat, for instance, the limit of nitrogenous manuring is determined less by what the plant can make use of than by what it can stand up to without lodging.

In practice it will seldom be possible to include every variety it is desired to test in the fertilizer trials. This, however, is no argument against including a representative selection of varieties. If no substantial differences in fertilizer response are discovered with such a selection we may then, reasoning inductively, conclude that it is improbable that substantial differences exist for *any* variety.

The method of experimentation in which two or more sets of treatments, or treatments and varieties, are taken in all combinations, was originally called complex experimentation, but inasmuch as this term may be taken to imply complexities of other kinds, I propose, following Fisher, to use the terms *factorial design* and *factorial experiments.*

3. *History.*

The idea of using all combinations of various sets of treatments in fertilizer experiments, and not only those which appear advisable on the grounds of the particular theory held at the moment, is a very old one. The two contrasting methods are well exemplified in classical fertilizer experiments on wheat and barley at Rothamsted.

The wheat experiment on Broadbalk was first laid down in the season 1843–4, but was first put on a permanent basis in 1851–2. Table I gives the table of treatments, the majority of which have been continued without change ever since. It will be seen that the effect of any particular mineral salt is for the most part given by

TABLE I.

Broadbalk Permanent Wheat.

Scheme of Manuring adopted in 1852.

Plot. No.	Manuring.				
2*b*	Farmyard manure				
3–4, 2*a* (half), 20	No manure				
5	—	P	K	Na	Mg
6	½N	P	K	Na	Mg
7, 15*a*	N	P	K	Na	Mg
8	1½N	P	K	Na	Mg
16	2N	P	K	Na	Mg
10	N	—	—	—	—
11	N	P	—	—	—
13	N	P	K	—	—
12	N	P	—	3⅔Na	—
14	N	P	—	—	2¾Mg
2*a* (half)	—	—	K	Na	Mg
17 } alter-	{ N	—	—	—	—
18 } nating	{ —	P	K	Na	Mg
9	Nitrate of Soda, etc.				
15*b*, 19	Rape Cake, etc.				

N = Ammonium salts.
P = Superphosphate.
K = Sulphate of potash.
Na = Sulphate of soda.
Mg = Sulphate of magnesia.

the difference of only a single pair of plots. On the other hand, the effect of increasing dressings of nitrogen is well determined by five levels of nitrogenous manuring. In plan the experiment consisted of a series of long, narrow plots (349 yards × 6·90 yards), each ½ acre in area, stretching the full length of the field. The numbering was consecutive from 2 to 19 across the field. A few of the plots were divided longitudinally into two halves (indicated by *a* and *b*). 3–4 was one plot. 2*a* was divided into two halves transversely, and 20 was one-third the length of the field.

The design of the Hoosfield barley experiment, laid out in 1852, is shown in Fig. 1. The mineral treatments were in strips along the field, and the nitrogen treatments in strips at right angles to them, so that all combinations of

$$\left\{\begin{array}{l}\text{No Superphosphate}\\ \text{Superphosphate}\end{array}\right\} \times \left\{\begin{array}{l}\text{No Potash}\\ \text{Potash, etc.}\end{array}\right\} \times \left\{\begin{array}{l}\text{No Nitrogen}\\ \text{Ammonium Salts}\\ \text{Nitrate of Soda}\\ \text{Rape Cake}\end{array}\right\}$$

were included, as well as farmyard manure and a few miscellaneous treatments.

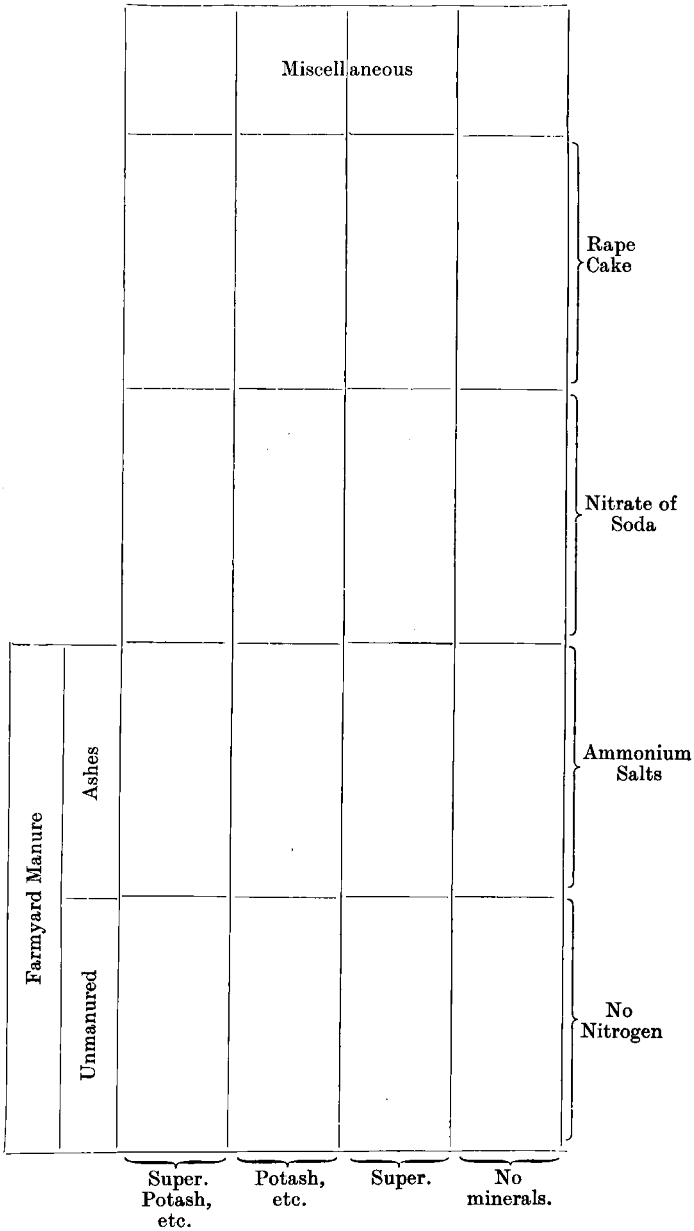

FIG. 1.

Hoosfield Permanent Barley.
Plan and Manures, 1852.

It will be seen that almost the whole of the field enters into the contrast between potash and no potash, and superphosphate and no superphosphate, and into the response to the various forms of nitrogen. Moreover, information is provided on the differences in response to any one of the fertilizers in the presence and absence of the others. Thus in one respect the design is markedly superior to Broadbalk. Unfortunately, however, the effect of broad fertility irregularities is likely to be serious. If, for instance, there is a fertility gradient from bottom to top of the field, even with no real effect of nitrogen a response to nitrogen would be indicated, rape cake being apparently the best. The fact that all four rape cake plots did better than all four no-nitrogen plots, which if the plots were arranged at random might be taken as evidence of a real difference, could here be equally regarded as evidence of fertility differences. In Broadbalk, on the other hand, fertility differences in one direction (actually up and down the slope) are entirely eliminated, since the plots stretch the whole length of the field. A certain amount of internal replication is also accidentally provided if it is assumed (as seems likely) that the effects of sulphate of soda and sulphate of magnesia are small.

The realization of the large and unknown errors introduced by fertility differences led to the use of replication, and also to a tendency to confine attention to a few simple comparisons, so that greater accuracy might be attained. Fertilizer experiments, for instance, tended to a simplified Broadbalk system of treatments, such as

$$(1),\ n,\ np,\ npk,$$

or

$$(1),\ np,\ nk,\ pk,\ npk,$$

the particular choice of treatments depending largely on the particular theory of fertilizer response held by the experimenter. (Throughout the paper the separate treatments will be indicated by small letters and the treatment effects and their interactions by capital letters. No treatment is indicated by (1), or by the suffix 0. The use of (1) instead of 0 for no treatment enables the expressions for the main effects and interactions to be written down by the rules of algebra.)

The development of complex experimentation in modern agricultural research is primarily due to Fisher. As early as 1926[2] he made a very strong recommendation in favour of complex experiments, which it will be of interest to recall.

> "In most experiments involving manuring or cultural treatment, the comparisons involving single factors, *e.g.* with or without phosphate, are of far higher interest and practical

importance than the much more numerous possible comparisons involving several factors. This circumstance, through a process of reasoning which can best be illustrated by a practical example, leads to the remarkable consequence that large and complex experiments have a much higher efficiency than simple ones. 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."

Since this date complex experimentation based on factorial design has been extensively practised at Rothamsted, and to a lesser extent elsewhere. The methods have evolved gradually, particularly in the direction of confounding, and with growing familiarity has come increased confidence. In the following sections I propose to give an outline of the methods at present in use. It is manifestly impossible, within the scope of a single paper, to enter into explanations of the detailed mechanism of all the devices. Detailed examples of three arrangements have, however, been included as exemplifying most of the points at issue.

4. *The* $2 \times 2 \times 2 \times \ldots$ *System of Treatments.*

The simplest type of factorial design is that in which there are only two treatments in each set, so that the total number of combinations is some power of 2. If we ignore the different forms of nitrogen the system of treatments in the Hoosfield barley experiment already referred to is of this type.

Let us first consider the case where there are two treatments, say nitrogen (n) and potash (k) only. We then have the four treatment combinations

$$(1),\ n,\ k,\ nk.$$

In the experiment on peas referred to in detail later the total yields on the six replicates of these treatments (ignoring slag, which in fact produced no effect) were respectively

(1)	n	k	nk
317·3,	365·1,	307·5,	327·1.

There are three independent comparisons, represented in the analysis of variance by three degrees of freedom. Clearly the comparisons can be made in many ways. We might, for instance, consider the response over the six replicates to n alone, given by

$n - (1)$, here 47·8, and to k alone, given by $k - (1)$, here $-$ 9·8, and the increase or decrease over the sum of these responses when n and k are applied in combination, namely $(nk - (1)) - (n - (1)) - (k - (1))$, or $nk - n - k + (1)$, here $-$ 28·2.

The difference between nk and k, however, also furnishes information on the response to n, being the response to n in the presence of k, and if the response to n and k are, in fact, independent, this response is equal to the response to n in the absence of k, except for experimental errors. A general measure of the effect of n is therefore provided by the mean of the responses to n in the presence and absence of k. The appropriate estimate of this is

$$N = \tfrac{1}{2}\{(nk - k) + (n - (1))\} = \tfrac{1}{2}(n - (1))\,(k + (1))$$
$$= \tfrac{1}{2}(19{\cdot}6 + 47{\cdot}8) = \tfrac{1}{2}(67{\cdot}4) = 33{\cdot}7,$$

and this may be defined as the *main effect* of n.

Equally the mean response to k in the presence and absence of n, *i.e.* the main effect of k, is estimated from

$$K = \tfrac{1}{2}\{(nk - n) + (k - (1))\} = \tfrac{1}{2}(n + (1))\,(k - (1))$$
$$= \tfrac{1}{2}(-38{\cdot}0 - 9{\cdot}8) = \tfrac{1}{2}(-47{\cdot}8) = -23{\cdot}9.$$

A measure of the lack of independence of n and k is given by the difference of the responses to n in the presence of k and in the absence of k, namely $(nk - k) - (n - (1))$, or what is the same thing, the difference of the responses to k in the presence and absence of n, namely $(nk - n) - (k - (1))$. One half of either of these is equal to

$$N \times K = \tfrac{1}{2}\{nk - n - k + (1)\} = \tfrac{1}{2}(n - (1))\,(k - (1))$$
$$= \tfrac{1}{2}(19{\cdot}6 - 47{\cdot}8) = \tfrac{1}{2}(-38{\cdot}0 + 9{\cdot}8) = \tfrac{1}{2}(-28{\cdot}2) = -14{\cdot}1.$$

This quantity is called the *interaction* between n and k, and represents one-half the differential response of n in the presence and absence of k.

The conventional introduction * of the factor $\tfrac{1}{2}$ has two advantages: first, the main effects and the interaction are all determined with equal precision, and secondly, the differences between individual treatment combinations can be determined by the direct addition or subtraction of the appropriate main effects and interactions. Thus the response to n in the absence of k is given by

$$N - (N \times K) = 33{\cdot}7 - (-14{\cdot}1) = 47{\cdot}8,$$

and the response to both fertilizers in combination, $nk - (1)$, is given by

$$N + K = 33{\cdot}7 - 23{\cdot}9 = 9{\cdot}8.$$

* This was made since the meeting.

It will be noted that this last response is not affected by the interaction.

It should be noted that the expressions for the main effects and interactions are really a matter of definition, the interactions being measures of the departure of the observed differences from the law implied in the definition of the main effects. Here the main effects are so defined as to imply an additive law between the effect of n and the effect of k; this is statistically convenient, and in agriculture appears to provide a good representation of the type of effects usually observed. But it should be clearly understood that the additive law has been provisionally imposed by the statistician and is not implicit in the data.

All the above responses are expressed in terms of the effects on totals of six plot yields. In practice the final presentation of the results will be in terms of some such units as hundredweight per acre, involving some further conversion factor. In the computation the factor $\frac{1}{2}$ will be combined with this conversion factor. Thus in the above example, where each plot has an area of 1/70th acre, and the yields are in pounds, the main effects and interaction will be found by multiplying 67·4, — 47·8, and — 28·2 (each of which is composed of the sums and differences of 24 plot yields) by

$$\frac{70}{12 \times 112}$$

to give the responses and interaction in hundredweight per acre.

It may be noted that the quantities 67·4, — 47·8, and — 28·2 are the differences of the marginal totals and the cross difference of the two-way table :

	(1).	n.	Total.
(1)	317·3	365·1	682·4
k	307·5	327·1	634·6
Total	624·8	692·2	1317·0

The sums of squares in the analysis of variance corresponding to these three quantities can be found by squaring the three numbers 67·4, — 47·8, and — 28·2 and dividing each square by 24, since each is the difference of two sums of 12 plots each. (The appropriate divisor of the square of any linear function of the plot yields

$$l_1y_1 + l_2y_2 + \ldots .$$

representing a single degree of freedom is $l_1^2 + l_2^2 + \ldots$. The same rule holds for finding the divisor of the sum of squares of deviations of quantities whose differences represent a set of two or more degrees of freedom, provided that no plot yield occurs in more than one quantity.)

The subdivision of the treatment sum of squares is given in Table II.

TABLE II.

	D.F.	Sum of Squares.	Mean Square.
Nitrogen	1	189·28	189·28
Potash	1	95·20	95·20
Interaction	1	33·14	33·14
Total	3	317·62	105·87

These three sums of squares possess an interesting statistical property, in that they add up to the total sum of squares (three degrees of freedom) for treatments. This is an indication that the three estimates are statistically independent. Technically the three linear functions corresponding to the three single degrees of freedom are said to be *orthogonal.* For a full discussion of the practical implications of orthogonality and non-orthogonality the reader is referred to Yates [8]. Two linear functions of the plot yields

$$l_1y_1 + l_2y_2 + \dots .$$
$$l_1'y_1 + l_2'y_2 + \dots .$$

are orthogonal if

$$l_1l_1' + l_2l_2' + \dots . = 0.$$

Although the appropriate estimates of the treatment effects are not always orthogonal, this is usually the case in a well-designed experiment, a fact which introduces a certain logical completeness into the analysis of variance. The additive property of sums of squares corresponding to orthogonal degrees of freedom is extensively used in the practical calculations required in the analysis of variance. It is important, therefore, that workers should be able to recognize the existence of non-orthogonality so as not to assume the additive property when it does not hold, an assumption which will lead to very serious errors.

The first two of the comparisons given in the third paragraph of this section are not orthogonal. The linear functions are

$$y_n + y_n' + y_n'' + \dots - y_0 - y_0' - y_0'' - \dots$$
$$y_k + y_k' + y_k'' + \dots - y_0 - y_0' - y_0'' - \dots$$

and the sum of the products ll' is $+ 6$. The sums of squares are

$\frac{1}{12}\{n - (1)\}^2$	190·40
$\frac{1}{12}\{k - (1)\}^2$	8·00
$\frac{1}{24}\{nk - n - k + (1)\}^2$	33·14
Total	231·54

The total is very different from the total sum of squares for treatments, 317·62. But each of these sums of squares may be legitimately compared with the error mean square by means of the z test, if it is desired to test the effect in question, just as the quantities $n - (1) = 47{\cdot}8$, etc. may be tested by the t test, which is the exact equivalent of the z test for a single degree of freedom.

The error mean square in this experiment (twelve degrees of freedom) is 15·44 (Table VI), so that the mean effects of both nitrogen and potash (Table II) are both significant (potash giving a depression), but the interaction is not significant.

It is, of course, possible that an interaction of some kind exists although it is not significant. It will be noticed that the response to n is only significant in the absence of k, and the depression with k is only significant in the presence of n. But in the absence of consistent interactions of this type over a series of experiments and of any knowledge of when they are likely to occur, the average response to n, 33·7, and the average depression with k, $-$ 23·9, are the best estimates to adopt in assessing the advantages of these fertilizers.

It will also be noticed that should the interaction exist and the yields of the individual combinations of treatments be substantially correct, the experimenter who divided his experiment into two parts, one on n with a basal dressing of k, and one on k with a basal dressing of n, would certainly discover the inadvisability of applying k, but would also come to the conclusion that the response to n was quite trivial, a seriously erroneous judgment.

In general if there is no evidence of interaction the mean responses to the two factors may be taken as the appropriate measures of the responses to these factors, which may be regarded as additive. If interaction exists, then usually information will be required on the responses to each factor in the presence and absence of the other. The experiment furnishes information on this point, but with only half the precision (*i.e.* with $\sqrt{2}$ times the standard error) of the mean comparisons. Nothing is lost, however, as even should the results be judged insufficiently precise, the information already obtained can be combined with that added later.

In certain cases it may be judged as a result of the first series of experiments, that more precise information will only be required on certain of the comparisons embodied in the factorial design. One variety, for instance, might prove itself so superior to all others that interest in response to fertilizers centred on the response of that variety. But the fact which may have emerged that some other variety did respond considerably more, although of no immediate practical interest, might still be of considerable value to the plant breeder in his task of combining the desirable features of several varieties.

The 2×2 system may be easily extended to any power of 2. We need only consider the case of a $2 \times 2 \times 2$ system. If the treatments are the three standard fertilizers n, p and k in all combinations, the main effect of n may be defined as the mean of the responses to n in the presence of all combinations of the other two fertilizers, and will then be estimated from

$$\begin{aligned} N &= \tfrac{1}{4}\{(n - (1)) + (np - p) + (nk - k) + (npk - pk)\} \\ &= \tfrac{1}{4}(n - (1))\,(p + (1))\,(k + (1)). \end{aligned}$$

The first order interactions of n and p may be defined as the mean of the interactions of n and p in the presence and in the absence of k, and will then be estimated from

$$\begin{aligned} N \times P &= \tfrac{1}{2}\{\tfrac{1}{2}(np - n - p + (1)) + \tfrac{1}{2}(npk - nk - pk + k)\} \\ &= \tfrac{1}{4}(n - (1))\,(p - (1))\,(k + (1)). \end{aligned}$$

The expressions for the main effects of p and of k, and the first order interactions of n and k and of p and k, will be similar. There is still one degree of freedom unaccounted for. This represents what is called the *second order interaction* between n, p and k, which is one half the difference between the interactions of n and p in the presence and absence of k, or between the interactions of n and k in the presence and absence of p, or between the interactions of p and k in the presence and absence of n, all these being equivalent and equal to

$$\begin{aligned} N \times P \times K &= \tfrac{1}{2}\{\tfrac{1}{2}(npk - nk - pk + k) - \tfrac{1}{2}(np - n - p + (1))\} \\ &= \tfrac{1}{4}(n - (1))\,(p - (1))\,(k - (1)). \end{aligned}$$

A numerical example of this type of subdivision is provided by the experiment on peas at Biggleswade, further discussed in Section 7.

The expressions tabulated on page 195 for various typical responses may be noted.

If the second order interaction is ignored the response to all three factors in conjunction is equal to the sum of the main effects of the three factors.

Response to:	Expression in terms of treatment combinations.	Expression in terms of main effects and interactions.
n (p absent, mean of k and no k)	$\frac{1}{2}\{nk + n - k - (1)\}$	$N - (N \times P)$
n (p and k absent)	$n - (1)$	$N - (N \times P) - (N \times K) + (N \times P \times K)$
n and p (mean of k and no k) ...	$\frac{1}{2}\{npk + np - k - (1)\}$	$N + P$
n and p (k absent)	$np - (1)$	$N + P - (N \times K) - (P \times K)$
n, p and k (complete fertilizer) ...	$npk - (1)$	$N + P + K + (N \times P \times K)$

5. *Subdivision of Sets of Degrees of Freedom.*

Very frequently we wish to introduce sets of treatments which contain more than two treatments each. In experiments involving fertilizers, for instance, it may be desirable to investigate more than one level of each fertilizer, or several different forms of the same factor, as in the nitrogen treatments of Hoosfield barley. In combined varietal and manurial trials it is usually advisable to include at least three varieties.

The primary subdivision of the treatment degrees of freedom follows the same lines as in $2 \times 2 \times 2 \times \ldots$ experiments, but each main effect and interaction will now contain more than a single degree of freedom.

The actual structure is best made clear by an example. Table III

TABLE III.

Oats Variety and Manuring Experiment.

Treatment Totals.

	n_0.	n_1.	n_2.	n_3.	Total.
v_1	429	538	665	711	2,343
v_2	480	591	688	749	2,508
v_3	520	651	703	761	2,635
Total... ...	1,429	1,780	2,056	2,221	7,486

gives the treatment totals (six replicates) of an experiment on oats involving three varieties and four levels of nitrogen. (The design of this experiment, which was a split-plot one, will be discussed in detail later.)

The partition of the treatment degrees of freedom and sum of squares is as follows:

	D.F.	Sum of Squares.	Mean Square.
Varieties	2	1786·36	893·18
Nitrogen	3	20020·50	6673·50
Interaction	6	321·75	53·63
Treatments... ...	11	22128·61	—

The sum of squares due to varieties is calculated from the varietal marginal totals, and that due to nitrogen from the nitrogen totals, while the interaction is usually obtained by subtraction of these two sums of squares from the total sum of squares due to treatments. The computation is exactly analogous to the computation of the sums of squares for blocks, treatments and error in a randomized block experiment. In fact "error" in such experiments is formally composed of the interactions of blocks and treatments.

This, however, does not exhaust the possibilities of subdivision. Any set of n degrees of freedom can be divided into n single degrees of freedom in an infinity of ways. Only the division which coincides with whatever physical facts it is desired to bring into prominence will normally be of interest, but in certain cases some especially simple formal division is useful for purposes of confounding.

In our example the three degrees of freedom for nitrogen may naturally be divided into the one degree of freedom representing the linear component of the response curve, and a second representing the quadratic component, and a third the cubic.

The dressings represent equal increments of nitrogen and the linear term is therefore (Fisher [1], Section 27) given by some fraction of

$$-3S(n_0) - S(n_1) + S(n_2) + 3S(n_3),$$

the quadratic term by

$$S(n_0) - S(n_1) - S(n_2) + S(n_3),$$

and the cubic term by

$$-S(n_0) + 3S(n_1) - 3S(n_2) + S(n_3).$$

The squares of these quantities, divided by 360, 72 and 360 respectively, will give the sums of squares attributable to each degree of freedom. These are

	D.F.	Sum of Squares.
Linear term	1	19536·4
Quadratic term	1	480·5
Cubic term	1	3·6
Total	3	20020·5

Most of the variation due to nitrogen response is accounted for by the linear term. The appropriate error mean square is 177·08, and the quadratic term is therefore not significant. The cubic term is below expectation.

The interactions between varieties and nitrogen may be similarly split up by calculating regression terms for each variety and taking the sums of the squares of the deviations between them. Indeed in testing for differential response this should be done, since such differential response is most likely to reveal itself in differences between the regressions for the different varieties; the whole set of interactions contains four degrees of freedom which are likely in any case to be small when there is little curvature in the average response curve.

In this particular example the total sum of squares for the six degrees of freedom is only 321·75, so that no interaction can possibly be significant, but the calculation may be performed as a formal exercise. The numerical values of the regression terms are

	Linear.	Quadratic.	Cubic.
v_1	973	63	99
v_2	904	50	22
v_3	775	73	−85

The sums of the squares of the deviations of these quantities divided by 120, 24 and 120 respectively give the following sums of squares :

	D.F.	Sum of Squares.	Mean Square.
$N_1 \times$ varieties	2	168·35	84·18
$N_2 \times$ varieties	2	11·08	5·54
$N_3 \times$ varieties	2	142·32	71·16
Total	6	321·75	—

If it were desired to test some theoretical response curve, the appropriate division of the sum of squares would be into the part accounted for by the curve, with as many degrees of freedom as there were arbitrary constants in the curve, and a part representing deviations from the curve. The test of the adequacy of the theory would be the non-significance of the latter response.

6. *Split Plot Arrangements.*

As already mentioned, the chief practical difficulty in the application of factorial design to agricultural field experiments is the fact that the number of treatment combinations rapidly becomes large with increasing complexity, with resultant large blocks and inadequate elimination of fertility differences. A further practical difficulty is that many agricultural operations cannot be conveniently carried out on small plots.

To meet these difficulties various modifications of the randomized block and Latin square lay-outs have been devised. The two of chief interest are the splitting of plots for subsidiary treatments, and the confounding of high-order interactions. We will first consider the split-plot type of design.

In principle this is very simple. The whole-plot treatments are arranged in the ordinary manner in a randomized block or Latin square, and each plot is subdivided into two or more parts to which are assigned at random the two or more sub-treatments. The analysis is also simple, especially if the division is only into two. In this case all that is necessary is to perform two separate analyses, with separate errors, one on the totals of the whole plots, and the

v_3	n_3 156	n_2 118	n_2 109	n_3 99	v_3
	n_1 140	n_0 105	n_0 63	n_1 70	
v_1	n_0 111	n_1 130	n_0 80	n_2 94	v_2
	n_3 174	n_2 157	n_3 126	n_1 82	
v_2	n_0 117	n_1 114	n_1 90	n_2 100	v_1
	n_2 161	n_3 141	n_3 116	n_0 62	
v_3	n_2 104	n_0 70	n_3 96	n_0 60	v_2
	n_1 89	n_3 117	n_2 89	n_1 102	
v_1	n_3 122	n_0 74	n_2 112	n_3 86	v_1
	n_1 89	n_2 81	n_0 68	n_1 64	
v_2	n_1 103	n_0 64	n_2 132	n_3 124	v_3
	n_2 132	n_3 133	n_1 129	n_0 89	
v_2	n_1 108	n_2 126	n_2 118	n_0 53	v_1
	n_3 149	n_0 70	n_3 113	n_1 74	
v_3	n_3 144	n_1 124	n_3 104	n_2 86	v_2
	n_2 121	n_0 96	n_0 89	n_1 82	
v_1	n_0 61	n_3 100	n_0 97	n_1 99	v_3
	n_1 91	n_2 97	n_2 119	n_3 121	

← rows →

Area of each plot : 1/80 acre. (28·4 links × 44 link rows.)

FIG. 2.

Oats Variety and Manuring Experiment.
Plan and yields in quarter lb.

other on the differences of the pairs of sub-plots. This latter analysis will contain an extra degree of freedom representing the direct effect of the sub-plot treatment. The interactions of the sub-plot treatment with the whole plot treatments will also appear in this analysis. If the subdivision is into more than two parts, the sums of squares corresponding to the components of the second analysis are usually obtained by subtraction, in the same manner as the interaction sum of squares in Section 5.

The oats variety and manurial experiment already referred to furnishes a good example of this type of lay-out. The varieties were sown in six randomized blocks of three plots each, and each plot was subdivided into four for the four levels of nitrogen.

The plan and yields of grain are shown in Fig. 2. The full analysis of variance is given in Table IV. The analysis is on a sub-plot basis, sums of squares from the whole-plot totals being divided by an extra 4, since each is the total of four sub-plots.

TABLE IV.

Oats Variety and Manuring Experiment.

Analysis of Variance (Sub-Plot Basis).

		D.F.	Sum of Squares.	Mean Square.
Whole Plots	Blocks	5	15875·28	3175·06
	Varieties	2	1786·36	893·18
	Error	10	6013·30	601·33
	Total	17	23674·94	—
Sub-Plots	Nitrogen	3	20020·50	6673·50
	N × Varieties... ...	6	321·75	53·63
	Error	45	7968·76	177·08
	Total	71	51985·95	—

In general in such experiments the comparisons between the sub-plot treatments are likely to be more accurately determined than those between the whole-plot treatments, since they depend on comparisons between closely adjacent plots. This is the case here, where the two error mean squares are 177·08 and 601·33, or in the ratio of 1 to 3·40. (The reader should satisfy himself that the ratio between these mean squares does represent the comparative accuracy of means containing the same number of sub-plots but subject to sub-plot and whole-plot errors respectively.) By replacing each treatment mean square by the corresponding error mean square we obtain the equivalent of a uniformity trial. Combining the two resultant error sums of squares then gives an estimate of what the error mean square would have been had all treatments combinations been randomized. This gives

$$\frac{12 \times 601 \cdot 33 + 54 \times 177 \cdot 08}{66} = 254 \cdot 22$$

subject to errors of estimation. Thus the accuracy of the varietal comparison would have been considerably increased, with some, but not a proportionate, loss of accuracy on the responses to nitrogen and their interactions with varieties. On the varietal comparisons we obtain 2·37 (=601·33/254·22) times the original information, and on the comparisons involving nitrogen 0·70 (= 177·08/254·22) times the original information.

In general, therefore, if the main effects of all treatments are required with equal accuracy, split-plot arrangements are not to be recommended where the whole-plot treatments are arranged in randomized blocks, except in the case in which whole-plot treatments are such that they could not be conveniently applied to such small plots as the sub-plot treatments.

If, however, the whole-plot treatments can be arranged in the form of a Latin square the situation is somewhat different, for if the whole of the treatments were randomized the Latin square arrangement might have to be sacrificed and a usually less efficient randomized block arrangement substituted.

In order to see what differences would result in practice by complete randomization in place of the use of split-plots, the split-plot experiments carried out at Rothamsted and its associated centres were examined. For the randomized block experiments new errors were computed as above. For Latin square experiments the mean of the sums of squares of rows and columns was first added into the whole-plot error, with a corresponding increase in degrees of freedom. Otherwise the procedure was the same as with randomized blocks.

The results are shown in Table V. In the case of randomized blocks there is a considerable gain in accuracy on the whole-plot comparisons, with, as must be the case, a corresponding loss of accuracy on the split-plot comparisons, relatively less in the case of splits into four. In the case of the Latin squares, however, with plots split into two, there is usually a loss of accuracy even on the main-plot comparisons owing to the transition from the Latin square to a randomized block arrangement. In the case of splits into four there is little to choose on the whole-plot comparisons but a definite gain on the split-plot comparisons.

It can, of course, be objected that the treatment of rows or columns as if they were blocks is unfair to the randomized block method, in that more compact blocks would, in fact, be chosen. The results given later in this paper, however, do indicate that Latin square arrangements are in practice markedly more efficient, and it is doubtful if the free choice of blocks would on the average have

TABLE V.

Split-Plot Arrangements.

Percentage information that would have been obtained with ordinary randomized block arrangements.

	Latin Squares.				Randomized Blocks.			
Plots split into:	2.		4.		2.		4.	
Comparisons involving:	Whole Plots.	Split Plots.	Whole Plots.	Split Plots.	Whole Plots.	Split Plots.	Whole Plots.	Split Plots.
Percentage.	Numbers of Experiments.							
0– 20	—	2	—	—	—	—	—	—
20– 40	2	5	1	—	—	3	—	—
40– 60	6	5	1	1	—	1	—	—
60– 80	4	3	—	3	—	2	—	2
80–100	4	4	2	4	—	1	—	1
100–120	3	3	—	1	1	—	—	—
120–140	1	—	5	—	1	—	—	—
140–160	1	—	—	—	1	—	—	—
160–180	1	—	—	—	4	—	—	—
180–200	—	—	—	—	—	—	1	—
200–220	—	—	—	—	—	—	1	—
220–240	—	—	—	—	—	—	1	—
Total	22		9		7		3	

resulted in any appreciable reduction in the errors below those calculated in the construction of the above table.

One point which must be borne in mind when considering the relative efficiency of randomized blocks and Latin squares is the number of error degrees of freedom. The six degrees of freedom for error provided by the 4×4 Latin square have long been recognized as inadequate, at least by Fisher. Something of the order of twelve error degrees of freedom would appear desirable in all experimental results which may afterwards have to be considered in conjunction with other material, unless the effects under investigation are large in comparison with their experimental errors. This seems to me to be the governing consideration in favour of the $2 \times 2 \times 2$ confounded arrangement to be described later, as opposed to the use of a 4×4 Latin square with split plots, which on a comparison of errors only would appear to be decidedly more efficient. Two 4×4 Latin squares with split plots, or a $3 \times 2 \times 2$ system of treatments arranged in a 6×6 square with split plots, are not open to this objection and are likely to be a very efficient arrangements; unfortunately, however, they require 64 and 72 plots respectively, which is frequently more than is practicable.

We may therefore conclude that the Latin square arrangement with split plots is a very efficient way of introducing an extra set (especially a pair) of treatments into a factorial system which can otherwise be arranged in the form of a Latin square, provided that it furnishes sufficient error degrees of freedom on the main comparisons.

The treatment of whole blocks of an ordinary experiment with different subsidiary treatments is a modification of the split-plot type of arrangement. Thus different blocks of a fertilizer experiment might be sown with different varieties. Here practically no information is obtained on the average effects of the block treatments, but their interactions with the principal treatments are determined with full accuracy, though it should be noted that differential response to the principal treatments in the different blocks will affect these interactions; this objection is discussed in Section 12. Such a procedure undoubtedly gives a wider inductive basis to the results and for this reason is to be recommended, though it is doubtful if there are many cases in which information on the subsidiary treatments is really not required. In the case of such subsidiary treatments as farmyard manure, however, it might be maintained that a one year's experiment is in any case almost worthless in assessing its value, but that a knowledge of the differences in response to fertilizers in its presence and absence is of considerable interest and importance.

It may be noted here that although whole blocks of a randomized block experiment may be given different treatments, the procedure is inadmissible with the rows or columns of a Latin square, unless the interactions of the principal and subsidiary treatments are to be disregarded, owing to the resultant non-orthogonality of the interactions with the columns or rows respectively. The point is dealt with elsewhere.[8] The rows themselves may be split, *e.g.* for varieties, but this, of course, increases the number of plots, unless varietal differences are ignored at harvesting, and thus is not a particularly efficient arrangement.

Another type of arrangement that stands condemned, but for a different reason, namely, because of a biased error, is the "semi-Latin square." In this arrangement there are twice (or more generally k times) as many rows as columns, and every treatment occurs once in each column and once in each pair of (or each k) rows, which are treated as a unit. It is only possible to make unbiased estimates of the appropriate errors when the treatments are grouped so as to give the equivalent of a split-plot arrangement.

7. *Confounding.*

Confounding is a method of reducing the block size by not completely replicating within each block. We have seen that in a factorial experiment the treatment comparisons are divisible into main effects and interactions of varying complexity. High-order interactions are usually non-existent, or at least small in magnitude compared with their experimental errors, and for this reason, if no other, of little practical interest. Certain components of these high-order interactions appearing in the analysis of variance table (though by no means all) are derived from the direct comparisons of equal groups of the various treatment combinations, all combinations being included. If smaller blocks are used, each containing only the combinations belonging to one group, the contrasts between the groups will coincide with block differences. No information, except a trivial amount derived from inter-block comparisons, is therefore available on the corresponding interactions, which are said to be confounded with blocks.

Provided that the other constituents of treatments in the analysis of variance table are orthogonal with the confounded interactions the remaining comparisons will not be formally affected by the confounding, being entirely intra-block. Their accuracy may be expected to be increased owing to reduction of error due to the smaller blocks.

A very simple example is provided by the $2 \times 2 \times 2$ system of treatments. If the experiment is arranged in blocks of four plots, half of which contain the treatments

$$(1),\ np,\ nk,\ pk,$$

and the other half the treatments

$$n,\ p,\ k,\ npk,$$

the second order interaction will be confounded.

It is easy to see that the main effects and first order interactions are unaffected by block differences. In each two positive and two negative treatment combinations occur in each block.

The experiment on peas at Biggleswade which has already been considered will form a useful example. The plan and yields are given in Fig. 3. The treatment totals are as follows:

Sub-blocks (*a*).				Sub-blocks (*b*).			
(1)	*np*	*nk*	*pk*	*n*	*p*	*k*	*npk*
154·3	173·8	164·0	151·5	191·3	163·0	156·0	163·1

The full analysis of variance is given in Table VI.

The components of the treatment sum of squares may be checked by computing the treatment sum of squares from the treatment totals and also the confounded interaction as if confounding did not exist. These calculations give 384·79 and 37·00 respectively, and the

216·1	*pk* 49·5	(1) 46·8	*n* 62·0	*k* 45·5	200·5
	np 62·8	*nk* 57·0	*npk* 48·8	*p* 44·2	
229·8	*n* 59·8	*k* 55·5	*np* 52·0	*nk* 49·8	202·1
	npk 58·5	*p* 56·0	(1) 51·5	*pk* 48·8	
243·1	*p* 62·8	*n* 69·5	*nk* 57·2	*pk* 53·2	225·4
	npk 55·8	*k* 55·0	*np* 59·0	(1) 56·0	

ROWS

Area of each plot: 1/70 acre. (31·75 links × 44·7 link rows.)

FIG. 3.

Experiment on Peas.

Plan and yields in lbs.

difference represents the sum of squares due to treatments in the analysis. The significant results of the experiment have already been discussed.

TABLE VI.

Experiment on Peas.

Analysis of Variance.

	D.F.	Sum of Squares.	Mean Square.
Blocks	5	343·30	68·66
N	1	189·28	—
P	1	8·40	—
K	1	95·20	
N × P	1	21·28	—
N × K	1	33·14	—
P × K	1	0·48	—
Error	12	185·28	15·44
Total	23	876·36	—

Instead of always confounding the same set of interactions it is possible to confound different sets in different sets of blocks.

Such a procedure is called *partial confounding*. Some information will then be obtained on all degrees of freedom, the loss of information being spread over several sets of degrees of freedom instead of being confined to one set.

In a $2 \times 2 \times 2$ system, for instance, each first order interaction might be confounded in a quarter of the blocks, and the second order interaction in the remaining quarter. The advisability or otherwise of such a course depends on the relative importance of the first and second order interactions.

A somewhat more complex example is provided by the $3 \times 3 \times 3$ system. Consider first the 3×3 table :

	a_1.	a_2.	a_3.
b_1	x_1	x_2	x_3
b_2	y_1	y_2	y_3
b_3	z_1	z_2	z_3

There are eight degrees of freedom. The two degrees for each main effect are given by the contrasts of the marginal totals. The four degrees of freedom for interactions are given by similar contrasts of diagonal totals, as follows.

	D.F.	Contrasts between
A main effects	2	$x_1 + y_1 + z_1$, $x_2 + y_2 + z_2$, $x_3 + y_3 + z_3$
B main effects	2	$x_1 + x_2 + x_3$, $y_1 + y_2 + y_3$, $z_1 + z_2 + z_3$
Interactions	2	$x_1 + y_2 + z_3$, $x_2 + y_3 + z_1$, $x_3 + y_1 + z_2$
	2	$x_1 + y_3 + z_2$, $x_2 + y_1 + z_3$, $x_3 + y_2 + z_1$

The degrees of freedom in a $3 \times 3 \times 3$ arrangement are capable of similar subdivision. There are eight degrees of freedom for the second order interactions, which can be split into four sets of two. Just as in a 3×3 table the interactions are given by the contrasts of the diagonal totals, so in a $3 \times 3 \times 3$ table the second order interactions are given by the contrasts of totals of diagonal planes in three dimensions. Thus one of the sets (I) is given by the contrast of all the 1's, all the 2's and all the 3's in the following table :

	a_1.			a_2.			a_3.		
	b_1.	b_2.	b_3.	b_1.	b_2.	b_3.	b_1.	b_2.	b_3.
c_1	1	2	3	3	1	2	2	3	1
c_2	3	1	2	2	3	1	1	2	3
c_3	2	3	1	1	2	3	3	1	2

A second (II) is given by interchanging a_2 and a_3 in this table, and the other two (III and IV) are obtained by interchanging c_2 and c_3 in the first two sets. If the 1's, 2's and 3's of the above table are placed in separate blocks the two degrees of freedom for second order interactions represented by this table will be confounded. An example of this type of confounding is given later in the paper.

Since in a $3 \times 3 \times 3$ system the split of the eight degrees of freedom into four sets of two is purely formal, all sets are of equal importance. Partial confounding should therefore be resorted to, different sets being confounded in different blocks. With 108 plots equal information will be obtained on all four sets, which introduces satisfactory simplicity into the statement of the results.

In the above examples the confounded degrees of freedom belong to the set of highest order interaction degrees of freedom in the ordinary subdivision of the treatment degrees of freedom. Such arrangements are only possible in a limited number of symmetrical systems. In a $3 \times 2 \times 2$ system, for instance, the treatment degrees of freedom are normally partitioned into :

		D.F.			D.F.
Main Effects	A	2	Interactions	A × B	2
	B	1		A × C	2
	C	1		B × C	1
				A × B × C	2

Only the main effects and the interaction $B \times C$ can be completely confounded. Any other division of a complete replication involves more than one of the above sets of degrees of freedom.

To avoid sacrificing all information on the possibly important $B \times C$ interaction the balanced arrangement given in Table VII

TABLE VII.

$3 \times 2 \times 2$ *arrangement.*

Replication :	I.						II.						III.					
Block :	I*a*.			I*b*.			II*a*.			II*b*.			III*a*.			III*b*.		
Treatment :	*a*.	*b*.	*c*.	*a*.	*b*.	*c*.	*a*.	*b*.	*c*.	*a*.	*b*.	*c*.	*a*.	*b*.	*c*.	*a*.	*b*.	*c*.
	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
	0	1	0	0	1	1	0	1	1	0	1	0	0	1	1	0	1	0
	1	0	0	1	0	1	1	0	1	1	0	0	1	0	0	1	0	1
	1	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	0
	2	0	0	2	0	1	2	0	0	2	0	1	2	0	1	2	0	0
	2	1	1	2	1	0	2	1	1	2	1	0	2	1	0	2	1	1

was devised. In this arrangement $B \times C$ is confounded as little as possible in each division. The balance attained by the three replications performs an important function and leads to a considerable simplification of the computations.

In this arrangement as little as $\frac{1}{9}$ of the information on $B \times C$ (1 D.F.) is lost, as compared with a similar unconfounded comparison. On the second order interactions $A \times B \times C$ (2 D.F.)

$\frac{4}{9}$ of the information is lost. It may be noted that $1 \times \frac{1}{9} + 2 \times \frac{4}{9} = 1$, corresponding to the one degree of freedom involved in the division of each replication into two parts. This relation only holds with the balanced arrangement.

Similar balanced arrangements have been devised for the $3 \times 3 \times 2$ system of treatments, and these can be extended to any system of the type 3×2^n and $3 \times 3 \times 2^n$.

Higher degrees of confounding than those exemplified in the above examples are sometimes advantageous, but the degree of confounding that can be indulged in without involving important interactions is very limited. In the 2^5 system, for instance, any single degree of freedom including the fourth order interaction can be confounded by a single division into blocks of 16 plots. If a triple division into blocks of 8 plots is required, at least two second order interactions and one third order of the type $A \times B \times C$, $A \times D \times E$ and $B \times C \times D \times E$ must be confounded; if the fourth order interaction is confounded a first order interaction or main effect is also necessarily involved. While in the case of blocks of four plots at least two first order interactions are necessarily confounded, a possible set being $A \times B$, $C \times D$, $A \times C \times E$, $A \times D \times E$, $B \times C \times E$, $B \times D \times E$, $A \times B \times C \times D$.

A high degree of confounding is therefore not likely to be of any value in agricultural research. In other branches of research, however, where whatever corresponds to block size is more strictly limited it may be of considerable utility. With five replications of a 2^5 system in blocks of four, for instance, a different pair of first order interactions, and their associated second and third order interactions, could be confounded in each replication. The loss of information would then be $\frac{1}{5}$ on all first order, $\frac{2}{5}$ on all second order, and $\frac{1}{5}$ on all third order interactions.

A variant on the standard systems which is of frequent occurrence, and which modifies the strictly formal analysis, both of ordinary factorial experiments and confounded experiments, is the occurrence of dummy treatments. If, for example, we are investigating three levels and three forms of nitrogenous manuring in all combinations, the three plots of each replication receiving no nitrogen are, in fact, identical. In confounded experiments this leads to the occurrence of plots receiving the same treatments in different blocks of the same replication, and these plots can, if desired, be used to furnish information on the differences between these blocks and therefore on apparently confounded degrees of freedom.

It should be mentioned here that although information is available on partially confounded degrees of freedom, its presentation in the form of a table of yields of individual treatment combinations necessitates a certain amount of extra computation. The difficulty does not arise in experiments with certain degrees of freedom totally confounded, nor in partially confounded experiments when we are content to present only the unconfounded effects.

8. *The Estimation of Error from High Order Interactions.*

Another difficulty arising from the large number of treatment combinations in an elaborate factorial experiment is that even two replications of each treatment combination may give a larger number of plots than are required to give the necessary accuracy on the main effects.

To meet this difficulty the device has been introduced of estimating the error from certain unconfounded high order interactions, which from previous experience can be confidently expected to be small in relation to experimental error.

A good example of this procedure is provided by the $3 \times 3 \times 3$ type of arrangement. This is a very suitable arrangement for investigating the responses to the three standard fertilizers, since evidence is obtained on the curvature of the response curves as well as their gradient, knowledge of the former being vital when decisions as to the optional dressing have to be made. A single complete replication demands 27 plots, which are as many as can be conveniently undertaken by most non-experimental farms which are interested in fertilizer trials. Were a minimum of 54 plots to be demanded many possible experiments would be ruled out.

(211) 2575	(121) 2599	(202) 2189
(120) 2472	(220) 2517	(020) 2093
(200) 2517	(022) 2411	(210) 2354
(002) 2403	(110) 2252	(111) 2268
(010) 2220	(212) 2381	(001) 1926
(021) 2252	(201) 2067	(122) 2152
(101) 2295	(102) 2021	(221) 2349
(112) 2362	(011) 1953	(012) 2025
(222) 2434	(000) 1989	(100) 2106

← rows →

Treatments: (211) indicates n_2, p_1, k_1, etc.
Area of each plot: 1/10 acre. (50 links $\times$ 200 link rows.)

FIG. 4.

Sugar Beet Experiment at Colwick.
Plan and yields of roots in lbs.

As an example of this type of experiment we may take the experiment on sugar beet at Colwick in 1934. The plan and yields of roots are shown in Fig. 4. The experiment is arranged in blocks of nine plots, so that one of the pairs of degrees of freedom for the second order interaction is allotted to blocks. The error is estimated from the remaining second order interactions and from all the first order interactions except the interactions of the regressions. From physical considerations and practical experience these may be expected to be small in relation to error in an experiment of ordinary accuracy. The partially confounded N regr. $\times$ P regr. $\times$ K regr. can be separated from error if desired, but this is unlikely to be of importance unless the effects of fertilizers are very marked.

The full analysis of variance is given in Table VIII. N regr. is computed from the square of $S(n_2) - S(n_0)$ divided by 18, N dev. by $S(n_2) - 2S(n_1) + S(n_0)$ divided by 54, and the first order interactions from the functions indicated by the tables :

N regr. × P regr.				N regr. × P dev.				N dev. × P regr.				N dev. × P dev.			
	$p_0.$	$p_1.$	$p_2.$		$p_0.$	$p_1.$	$p_2.$		$p_0.$	$p_1.$	$p_2.$		$p_0.$	$p_1.$	$p_2.$
n_0	+1	0	−1	n_0	−1	+2	−1	n_0	−1	0	+1	n_0	+1	−2	+1
n_1	0	0	0	n_1	0	0	0	n_1	+2	0	−2	n_1	−2	+4	−2
n_2	−1	0	+1	n_2	+1	−2	+1	n_2	−1	0	+1	n_2	+1	−2	+1
Divisor: 12				36				36				108			

The construction of these tables is obvious if we remember that the linear response to n for p_0 and the mean of all k is given by $S(n_2p_0) - S(n_0p_0)$, with similar expressions for the linear response to n for p_1 and p_2. The linear regression of these three quantities is $-$ 1 times the first, 0 times the second, and $+$ 1 times the third, giving the first table; while the curvature is $+$ 1 times the first and third, and $-$ 2 times the second, giving the second table. The third and fourth tables are derived in a similar manner.

The numerical totals over all levels of k are :

	$p_0.$	$p_1.$	$p_2.$	Total.
n_0	6,318	6,198	6,756	19,272
n_1	6,422	6,882	7,223	20,527
n_2	6,773	7,310	7,300	21,383
Total	19,513	20,390	21,279	61,182

from which the n and p effects and their interactions can be calculated.

The division of the second order interactions has already been described.

TABLE VIII.

Sugar Beet Experiment at Colwick.

Analysis of Variance.

	D.F.	Sum of Squares.	Mean Square.
Blocks (I)	2	244,526	122,263
N regr.	1	247,573	—
N dev.	1	2,948	—
P regr.	1	173,264	—
P dev.	1	3	—
K regr.	1	1,120	—
K dev.	1	2,017	—
N regr. × P regr.	1	660	—
N regr. × K regr.	1	70,687	—
P regr. × K regr.	1	616	—
Error	15	262,298	17,487
Total	26	1,005,712	—

Components of Error.

	D.F.	Sum of Squares.		D.F.	Sum of Squares.
N regr × P dev.	1	41,684	P regr. × K dev.	1	26,136
N dev. × P regr.	1	11,271	P dev. × K regr.	1	28
N dev. × P dev.	1	1,261	P dev. × K dev.	1	972
N regr. × K dev.	1	6,110	N × P × K II	2	388
N dev. × K regr.	1	15,335	N × P × K III	2	5,742
N dev. × K dev.	1	95,230	N × P × K IV	2	58,141

In this experiment the response to both n and p reach the 1 per cent. level of significance, but in neither case is there any evidence of falling off in response with higher dressing nor is there any interaction between the two fertilizers. k shows no significant effects.

9. *Independent Factors.*

The estimation of error from high-order interactions leads naturally to the consideration of the case where all interactions can be assumed to be non-existent. A very simple example of this is provided by the operation of weighing.

Suppose that the weights of seven very light objects require to be determined, and that the apparatus being used necessitates an additional observation for zero correction. Suppose, further, that systematic errors are non-existent, or at least small compared with random errors. The obvious procedure would be to weigh each object separately, and to make an eighth weighing with no object to determine the zero correction. The efficiency, however, can be quadrupled by weighing the objects in groups according to the following scheme :

Weighing No.	Objects weighed.
1	$a + b + c + d + e + f + g$
2	$a + b \quad + d$
3	$a \quad + c \quad + e$
4	$a \quad + f + g$
5	$b + c \quad + f$
6	$b \quad + e \quad + g$
7	$c + d \quad + g$
8	$d + e + f$

In this scheme it will be noticed that every object is weighed four times, and that in the four weighings of a given object every other object is included twice, the remaining four weighings also including every other object twice. Thus object a is included in weighings 1, 2, 3, and 4, which together contain b, c, d, e, f, g twice each. Weighings 5, 6, 7, and 8 contain objects b, c, d, e, f, g twice each, but not a. The difference between the mean of 1, 2, 3 and 4, and the mean of 5, 6, 7 and 8, therefore provides an estimate of the weight of a, and since it is the difference of two means of four weighings the estimate has four times the precision (one-half the standard error) of that given by the ordinary procedure.

The formal analogy of the above scheme and the $2 \times 2 \times 2$ factorial system may be drawn here. If the weighings 1 to 8 are replaced by the treatment combinations *npk*, *np*, *nk*, *n*, *pk*, *p*, *k* and (1) respectively, it will be found that the estimate of the weight of a is transformed into the estimate of the main effect N. Similarly, the estimates of the weights of b, c, d, e, f and g are transformed respectively into the estimates of the main effects P and K, and the interactions $N \times P$, $N \times K$, $P \times K$ and $N \times P \times K$.

Cases in which the interactions are certainly negligible are, in fact, rather rare. Even in the example given above, although weights are undoubtedly additive, systematic errors of the apparatus are likely to complicate the issue. But such experimental systems may be useful in certain preliminary surveys, where there is good reason to believe that interactions are small and where it is desired to include as many factors as possible (some of them perhaps unlikely to produce any effect whatsoever).

The experiments described by Tippett [6] and already referred to in the introduction are of this type. He there considers designs for experiments containing 3, 4 or 5 factors, with five values for each factor, the designs being based on 5×5 Latin, Græco-Latin, and hyper-Græco Latin squares. Not only are the error degrees of freedom composed of interactions of the various factors, but the mean values representing the main effects are not in reality pure main effects, but also contain certain interaction components.

10. *Loss of Efficiency due to Increase in Block Size.*

As far as I know no very extensive investigation has been made of the increase in experimental error due to increased size of block, but Wishart [7] in a paper presented to the Empire Cotton-Growing Corporation Conference in 1934 advised against complex experiments on these grounds. He there concludes :

> " I know that there has been a tendency to proceed to complex experiments with two or more sets of interacting treatments, owing to the flexibility of the arrangement. The danger of loss of efficiency on this ground [the large size of the blocks], and on the ground of inadequate replication, should be guarded against by every experimenter."

The question of inadequate replication has already been dealt with. It has been shown that far from there being any loss of efficiency in factorial designs there is a very considerable gain. It remains to consider how far increase in plot error due to increase in block size is likely to outweigh the very real advantages of factorial design.

The question might be best approached by considering the efficiency of various types of arrangement, and in particular of varying block size, over a series of uniformity trials. This, however, has not been possible up to the present owing to the large amount of work involved.

Some indication of the relative merits of various types of arrangement can be gained from the examination of the results of experiments already carried out. Very little extra work is entailed, since the analyses are already in existence.

A comparison of the efficiency of split plots with that of complete randomization within blocks has already been described in Section 6. Two further comparisons have been made on Rothamsted material, one to determine the loss of efficiency that would result if experiments actually arranged in Latin squares or randomized blocks had been completely randomized, and the second to determine the loss of efficiency that would result if experiments actually confounded had not been confounded.

To investigate the effects of complete randomization it is necessary to make an estimate of what the error mean square would have been had there been no restrictions.

Suppose that the mean squares and degrees of freedom in the analyses of variance are as follows :

Randomized Blocks.			Latin Square.		
	D.F.	Mean Square.		D.F.	Mean Square.
Blocks ...	$q-1$	B	Rows ...	$p-1$	R
Treatments ...	$p-1$	T	Columns ...	$p-1$	C
Error ...	$(p-1)(q-1)$	E	Treatments...	$p-1$	T
			Error ...	$(p-1)(p-2)$	E

If the treatments had been dummy the treatment mean square would equal the error mean square, except for errors of estimation. Replacing T by E the total sum of squares in the case of the randomized block experiment would become

$$(q-1)B + q(p-1)E$$

with $pq - 1$ degree of freedom. If there were complete randomization the error mean square with dummy treatments would be derived directly from this. The relative efficiency is therefore given by the ratio

$$\frac{(pq-1)E}{(q-1)B + q(p-1)E}.$$

In the case of the Latin square similar reasoning gives the ratio

$$\frac{(p^2-1)E}{(p-1)R + (p-1)C + (p-1)^2E} = \frac{(p+1)E}{R + C + (p-1)E}.$$

The distribution of the percentage efficiencies for all experiments carried out in 1932 and 1933 is given in Table IX. The mean percentage efficiencies for the two years were as follows. (The numbers in brackets indicate the numbers of experiments.)

	1932.	1933.
Randomized Blocks	72·3 (22)	75·2 (22)
Latin Squares	54·1 (38)	57·4 (37)

The differences between the randomized block and Latin square arrangements are quite marked, especially with the larger sizes of blocks. Whereas the randomized block arrangements have only removed on the average 26 per cent. of whatever variation existed over the experimental site, Latin squares have accounted for no less than 44 per cent.

On the basis of these figures, therefore, one might expect a randomized block arrangement to have on an average an efficiency of only $\frac{56}{74}$, or about 75 per cent. of a Latin square on the same plots.

TABLE IX.

Percentage Efficiency with Complete Randomization.

	Latin Squares.			Randomized Blocks.		
	4 × 4.	5 × 5.	6 × 6 and 8 × 8.	4–6 treatments.	7–9 treatments.	10–24 treatments.
Percentage Efficiency.	Number of Experiments.					
0– 10	—	—	—	1	—	—
10– 20	5	2	—	—	—	—
20– 30	3	1	1	2	—	—
30– 40	8	7	—	1	—	1
40– 50	6	3	1	2	—	—
50– 60	4	2	1	2	1	—
60– 70	2	2	—	1	6	1
70– 80	5	4	—	4	1	1
80– 90	7	2	1	—	3	3
90–100	2	1	—	2	4	1
100–110	2	—	1	2	3	2
110–120	2	—	—	—	—	—
Total	46	24	5	17	18	9

A comparison of the average percentage standard errors given in Table X reveals an even greater advantage in favour of the Latin square. (Rothamsted and Woburn experiments have been excluded from this table, since in the later years they consisted of large factorial experiments not comparable with those of the outside

TABLE X.

Average Percentage Standard Errors per Plot.

Outside Centres.

	1927–30.	1931.	1932.	1933.	Weighted Mean.
Potatoes :—					
Randomized Blocks ...	9·0 (6)	10·2 (6)	4·5 (3)	10·6 (7)	9·2 (22)
Latin Squares	5·2 (15)	6·6 (11)	7·8 (15)	7·4 (15)	6·8 (56)
Sugar Beet, Roots :—					
Randomized Blocks ...	7·4 (5)	5·8 (2)	8·2 (7)	12·9 (1)	7·9 (15)
Latin Squares	6·4 (8)	5·3 (5)	7·1 (7)	5·5 (8)	6·1 (28)

centres.) In the case of potatoes the relative efficiency as judged by the ratio of the squares of the mean errors is 55 per cent., and with sugar beet it is 60 per cent. It is to be expected that the apparent efficiency of randomized blocks as compared with Latin squares would be found to be lower when judged on this comparison, as many of the Latin squares were only 4 × 4, so that presumably there was less fertility variation to remove than in the case of the larger randomized block experiments.

On the basis of these figures we can make a rough estimate of the relative efficiency of a factorial design and the equivalent simple arrangements. As an example we will consider an experiment on three levels of each of the three standard fertilizers. With 54 plots we should have the option of making three separate experiments of 18 plots each, or of carrying out a $3 \times 3 \times 3$ factorial experiment in six blocks of 9 plots (two replicates). The separate experiments might well be laid out in two 3×3 Latin squares each (six squares in all) with a pooled estimate of error (eighteen degrees of freedom) from the three experiments, provided these were all in the same field. Taking the error variance per plot in the Latin square arrangement as 50 per cent. of that in the factorial arrangement (*i.e.* somewhat less than the values of 55 per cent. and 60 per cent. given in the last paragraph), the relative efficiency on the main effects will be in the ratio of 3 : 2 in favour of the factorial design (since three times as many plots are involved in all means). Thus in mere accuracy of estimation of main effects the factorial design is 50 per cent. more efficient, as well as having the additional and even more important advantage of providing information on the interactions and a wider inductive basis for the results.

It will be noted that the factorial arrangement might be expected to be even more efficient relatively to the simple arrangements if these latter had been arranged in randomized blocks instead of Latin squares. It is not so much the increase in size of block in a complex experiment that causes a higher error variance per plot as the necessity of abandoning the admittedly more efficient Latin square arrangement. Indeed, as we have seen, that complete randomization without restrictions of the 1932 and 1933 randomized block experiments would only have decreased the efficiency by about 25 per cent. on the average.

The examination of the confounded experiments confirms this. The reduction in block size due to the confounding has resulted on the average in quite moderate increases in efficiency.

In making the estimation of loss of efficiency due to failure to confound it must be assumed that had the experiment not been confounded the blocks composing each complete replication would have formed one large block; it is, of course, possible that the experimenter might have laid out the experiment with differently shaped large blocks had he not been confounding. To simplify the calculations it has also been assumed that partially confounded interactions are negligible except where they have been judged significant. Completely confounded interactions have always been assumed negligible.

The method of calculation is as follows. Suppose that each block contains k plots, and that there are hk treatment combinations in all, so that in each replication $h - 1$ degrees of freedom are confounded. Let there be r replicates. The analysis of variance with a single set of degrees of freedom completely confounded is as follows.

		D.F.	Mean Square.
Blocks	Between complete replications... ...	$r - 1$	A
	Within " "	$r(h - 1)$	B
Treatments...	... ...	$h(k - 1)$	T
Error	... ...	$h(r - 1)(k - 1)$	E

Replacing T by E and including the second part of blocks gives the sum of squares

$$r(h - 1)B + rh(k - 1)E,$$

with $r(hk - 1)$ degrees of freedom, so that the percentage efficiency without confounding is

$$\frac{(hk - 1)E}{(h - 1)B + h(k - 1)E}.$$

Table XI gives the distribution of this fraction for various groups of Rothamsted experiments. In less than 10 per cent. of

TABLE XI.

Relative Efficiency without Confounding.

	2 × 2 × 2.		3 × 3 × 3.		Miscellaneous.	Total.
	Small Plots.	1/10 acre Plots.	Small Plots.	1/10 and 1/20 acre Plots.		
Percentage Efficiency.	Number of Experiments.					
0– 10	—	—	—	—	—	—
10– 20	—	—	—	—	—	—
20– 30	—	1	—	—	—	1
30– 40	1	—	1	—	—	2
40– 50	—	1	1	—	—	2
50– 60	—	—	3	—	2	5
60– 70	—	4	1	1	2	8
70– 80	4	1	5	2	2	14
80– 90	2	3	1	—	3	9
90–100	2	1	7	2	1	13
100–110	2	—	2	4	—	8
110–120	2	2	—	—	—	4
Total	13	13	21	9	10	66

the experiments has the efficiency been more than doubled by the confounding, but in over half the experiments it has been raised by more than 25 per cent. In about a third of the experiments the gain was either trivial or non-existent.

11. *A Further Consideration Governing Optimal Block Size.*

There is another aspect of the question of the most efficient block size which is of importance. It is often tacitly assumed that the block size which gives the minimal error variance per plot will be the most efficient, but in fact it is necessary to take into consideration not only the variation of the error with change in block size, but also the number of treatments it is desired to test.

The points involved can be brought into prominence by considering the relative efficiency of the (randomized) half-drill strip method and the randomized block or Latin square method of comparing several treatments. The discussion applies equally to any system of comparing the treatments in pairs.

Usually in the method of pairs one treatment is chosen as control and all others are compared with it. This has one obvious disadvantage, in that the comparisons with control are made with twice the accuracy (*i.e.* $1/\sqrt{2}$ of the standard error) of comparisons between any other pair of treatments, since if A is the control, the difference of B and C must be deduced from $\delta_m(B - A) - \delta_m(C - A)$, where $\delta_m(B - A)$ is the mean difference of B and A in the pairs where they occur together. This disadvantage can be obviated by making comparisons between every pair of treatments. Thus with five treatments, A, B, C, D, E, there are ten possible comparisons AB, AC, AD, AE, BC, BD, BE, CD, CE, DE. It is a remarkable fact that with this type of design, using the same number of plots, the same accuracy is obtained on all comparisons as would be obtained on the comparisons of the control and another treatment in the customary type of arrangement.

With the symmetrical arrangement, if σ'^2 is the error variance per plot, p the number of treatments, and n the available number of plots, the variance of the difference between two treatment means is

$$\frac{4(p-1)}{n}\sigma'^2.$$

With the " control " arrangement this is the variance of the $(p - 1)$ comparisons involving the control, while the other $\frac{1}{2}(p-1)$ $(p - 2)$ comparisons have double the variance. The mean variance of all comparisons is therefore

$$\frac{8(p-1)^2}{pn}\sigma'^2.$$

With an arrangement in randomized blocks or Latin squares, with an error variance of σ^2 per plot, the variance between two treatment means is

$$\frac{2p}{n}\sigma^2.$$

Thus, provided that σ^2/σ'^2 is less than $2(p-1)/p$, the randomized block or Latin square arrangement is more efficient than the symmetrical arrangement, while σ^2/σ'^2 must be greater than $4(p-1)^2/p^2$ for the control method to be the more efficient. With six treatments these fractions have the values of 5/3 and 25/9 respectively.

In agricultural field trials the greater similarity of pairs of plots over those of blocks of five is seldom likely to give a ratio of σ^2/σ'^2 as great as 25/9, or even 5/3.

Tables XII and XIII show the results of a test on two uniformity trials. In Mercer and Hall's wheat data [5] randomized blocks gave

TABLE XII.

Mercer and Hall. Wheat.

Plots used: 5 unit-plots along rows (E and W) × 1 unit-plot across rows, giving 100 plots (5 × 20) of 1/100 acre each.
Blocks of 5 plots across rows, or four 5 × 5 Latin squares.

	Error mean square.
Randomized blocks	0·940
Latin squares	0·723
Half-drill strip method	0·698
Randomized pairs	0·742

TABLE XIII.

Immer. Sugar Beet.

Plots used: 4 rods long × 4 rows wide, 2 edge rows rejected.
Rows 7–54 only used: 60 plots (5 × 12) of 1/90 acre each.
Blocks of 6 plots across rows, or two incomplete 6 × 6 Latin squares.

	Error mean square.
Randomized blocks	237·2
Latin squares	114·4
Half-drill strip method	130·6
Randomized pairs	161·8

an error mean square of 0·940, compared with 0·698 by the half-drill strip method, their ratio being 1·35. The method of randomized blocks is therefore much more efficient.

In Immer's sugar beet data [4] randomized blocks gave 237·2 and the half-drill strip method 130·6 with a ratio of 1·82. This is slightly greater than 5/3, but much less than 25/9. The symmetrical method would therefore be slightly more efficient than randomized blocks, but the control method considerably less so.

In both cases the Latin square arrangements have reduced the error to that of the half-drill strip method, so that the latter method must be judged very inferior to the use of Latin squares. In one case only has the half-drill strip method any appreciable advantage over random pairs. In fairness, however, it should be pointed out that the plots are wider than they would properly be on the half-drill strip method.

In other experimental material much greater similarity may exist between pairs than between larger blocks, as, for example, in human beings, where the similarity between monozygotic twins is strikingly greater than between any other groups that can be formed. In such cases the arrangement in symmetrical pairs may be of considerably greater efficiency than any arrangement based on the use of larger blocks.

It may be mentioned here that the method of symmetrical pairs is capable of extension to blocks of any number less than the number of treatments. Thus if the experimental material naturally forms blocks of three, six treatments may be tested by an arrangement such that (1) every treatment occurs equally frequently and not more than once in each block, (2) every pair of treatments occur together in the same number of blocks. For this type of arrangement I propose the name of *symmetrical incomplete randomized blocks.* Certain modifications are required in the analysis of variance, and in the presentation of results, but these modifications are surprisingly simple and demand very little extra computational labour.

12. *Differential Responses in Different Blocks.*

An objection that has sometimes been made against confounding is that the effects of one or more of the treatments may vary from block to block. This will tend to inflate some other interaction sum of squares. In the $2 \times 2 \times 2$ arrangement, for instance, if blocks containing the *npk* plots respond better to nitrogen than those containing no fertilizer it is easy to see that an apparent positive $P \times K$ interaction will result. There will, of course, be no bias in this interaction, for positive and negative values will occur with equal frequency, but it will be judged significant more often than it should if compared with the ordinary error.

The objection is not really one against confounding, but against the whole system of pooling estimates of error in the analysis of variance. If such variation in response exists, then the error is no longer homogeneous, and should be split up into its component parts, namely, the interaction of each treatment degree of freedom with blocks. The response to n, for instance, would then be tested against the variation of this response from block to block, namely

$N \times$ blocks, and in so far as the blocks could be regarded as a random sample of all possible blocks in the field, this would answer the question as to whether the response to nitrogen over the field could be regarded as significant or whether it was due to a lucky selection of blocks. This subdivision of errors cannot be made in Latin square arrangements.

In practice the force of the objection depends entirely on whether, in fact, there is likely to be any substantial variation in response from block to block in experiments ordinarily carried out. It is well known that response to fertilizers varies very greatly from farm to farm, and it would therefore appear possible that such variations occur even on different parts of the same field.

There are various methods of testing whether, in fact, this is so. One is to tabulate the mean squares of the interactions of blocks with each treatment degree of freedom, and see if they are homogeneous by testing against χ^2. Again, in confounded experiments the interaction of some particular main effect, say N, with sub-blocks can be compared with the remaining error degrees of freedom. Yet a third way, when there are split plot experiments available, is to examine the interaction of the split treatment with blocks or with the rows and columns of a Latin square.

This last test has been carried out on the experiments with split plots at Rothamsted and its associated centres. Only those experiments in which the split treatment shows a significant effect were chosen. In each case the probability of getting as large or larger apparent effect was calculated from the z distribution. Table XIV shows the results. In Latin squares rows and columns have been taken together if the error degrees of freedom were less than 40.

TABLE XIV.

Interaction with Soil Differences.

Year.	Place.	Crop.	Treatment.	Interaction with	n_1.	n_2.	P.
1929	Rothamsted (two squares for single and double nitrogen)	Barley	Potash	Rows	4	52	0·884
				Columns	4	52	0·878
				Rows	4	52	0·340
				Columns	4	52	0·752
1931	Rothamsted	Wheat	Harrowing	Blocks	3	24	0·157
,,	,,	Oats	Nitrogen	Blocks	5	40	0·614
,,	Badminton	Hay	Potash	Rows and Cols.	8	12	0·519
,,	Chesterfield	Hay	Potash	Rows and Cols.	8	12	0·415
,,	Downham	Potatoes	Nitrogen	Blocks	3	24	0·424
,,	,,	Sugar Beet	Nitrogen	Blocks	3	24	0·286
,,	Potton	Potatoes	N/S–S/A	Rows and Cols.	6	6	0·254
,,	,,	Cabbages	N/S–S/A	Rows and Cols.	6	6	0·245
1932	Rothamsted	Potatoes	Nitrogen	Blocks	2	94	0·155
,,	Colchester	Sugar Beet	Salt	Blocks	3	24	0·020
,,	Oakerthorpe	Mangolds	Dung	Rows and Cols.	6	6	0·349
1933	Rothamsted	Wheat	Nitrogen	Rows	3	41	0·530
				Columns	3	41	0·640
,,	,,	Forage	Nitrogen	Rows and Cols.	8	12	0·116
,,	Elsham	Sugar Beet	Nitrogen	Rows and Cols.	8	12	0·896

The table shows no evidence of any variation in response. There is only one significant result ($P = 0{\cdot}0195$), that of salt in the 1932 experiment at Colchester, and examination of the yields of the individual plots of this experiment did not support the theory of differential response.

The distribution of the probabilities is shown in Fig. 5. The

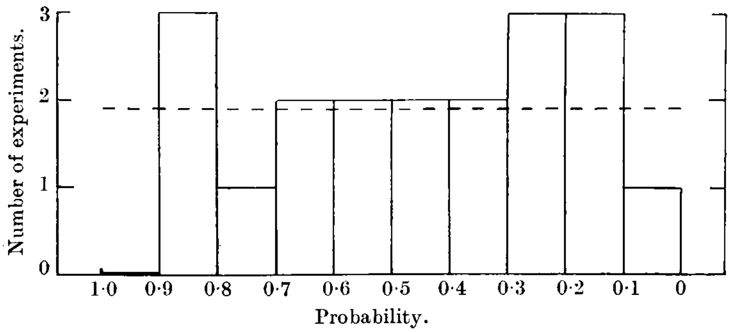

FIG. 5.

Interactions of treatments with soil differences: distribution of probabilities.

The dotted line shows the expectation in each class.

probability of obtaining a set of probabilities lower than these by chance is given by the sum of half the Napierian logarithms of the probabilities, which will be distributed as χ^2 with 38 degrees of freedom.[1] This probability was found to be 0·34.

13. *Conclusions.*

I hope that the results presented in this paper will be a sufficient demonstration that, rightly used, the complex methods of experimentation included in the term factorial design are of considerable value in agronomic research. The special devices of confounding and the estimation of error from high-order interactions admittedly complicate the statistical analysis of the results, but many of the difficulties vanish with familiarity, and others (such as the somewhat laborious calculations required to furnish a table of yields of all treatment combinations freed from block effects) arise from the attempt to present the results in a form which is excellent for an ordinary randomized block or Latin square experiment but unsuitable for a confounded arrangement.

There remains the danger of misinterpretation owing to faulty statistical analysis, and the perhaps more serious danger of faulty design. Both are real dangers, but the adoption of standard arrangements would appear to be an adequate safeguard. In practice most workers are unable to give special attention to the design and analysis of each individual experiment, and if only for this reason the use of standard patterns appears to be a necessity.

I have not attempted to discuss the relative practical difficulties of factorial and simple experiments. Many of these difficulties undoubtedly arise from unfamiliarity and vanish completely as soon as one or two factorial experiments are actually carried out. Of those that do not, the man actually responsible for the conduct of the experiments in the field must be the judge. At the same time, in order to judge fairly he must be able to assess the advantages of factorial design.

Factorial experiments, particularly confounded experiments, are more vulnerable than simple experiments with small blocks. The failure of any considerable part of a single block involves the results of the whole of that replication in which it occurs, and any use of the replication necessitates tiresome and lengthy computations. The failure of animals in animal husbandry experiments can be equally troublesome. For this reason excessively complex factorial experiments should not be undertaken when the experiments are exposed to serious natural hazards.

An important advantage of the straightforward randomized block arrangement is that an estimate of error can be isolated for every comparison separately. This is of very great value when handling new and unknown material, or treatments which may produce large differences and even partial or complete failures. In such cases the assumption of constancy of error variance is entirely unjustified, but in a randomized block experiment any treatment or treatments may be excluded and the analysis carried out on the remainder. This is not true of either the Latin square or of confounded arrangements.

I have already expressed the opinion that factorial design is likely to be of interest to workers in other fields of research, and that in many of these fields the complication of confounding is likely to be unnecessary. At Rothamsted we are using factorial designs in experiments with pigs, and also in pot culture work. I have not space to say more of these applications here, but I would like to conclude by expressing the hope that workers in other fields may find these methods as effective as they have already proved in agricultural research.

References.

[1] Fisher, R. A. *Statistical Methods for Research Workers* (1925). Edinburgh: Oliver and Boyd (5th Edition, 1935).

[2] Fisher, R. A. The arrangement of field experiments. *Jour. Min. Agric.* (1926), vol. 33, pp. 503–513.

[3] Fisher, R. A., and Wishart, J. The arrangement of field experiments and the statistical reduction of the results. Imperial Bureau of Soil Science, Technical Communication (1930), No. 10.

[4] Immer, F. R. Size and shape of plot in relation to field experiments with sugar beets. *Jour. Agric. Res.* (1932), vol. 44, pp. 649–668.
[5] Mercer, W. B., and Hall, A. D. The experimental error of field trials. *Jour. Agric. Sci.* (1911–12), vol. 4, pp. 107–132.
[6] Tippett, L. H. C. Some applications of statistical methods to the study of variation of quality of cotton yarn. *Jour. R.S.S. Supplement* (1935), vol. 2, pp. 27–62.
[7] Wishart, J. Analysis of variance and analysis of co-variance, their meaning, and their application in crop experimentation. Empire Cotton-Growing Corporation Conference, July 1934. Report, pp. 83–96.
[8] Yates, F. The principles of orthogonality and confounding in replicated experiments. *Jour. Agric. Sci.* (1933), vol. 23, pp. 108–145.

Discussion on Mr. Yates's Paper.

Sir William Dampier: I have two functions to perform to-day; one is to fill—however unworthily—this gorgeous Chair, and the other is to propose a cordial vote of thanks to Mr. Yates for the very interesting work which he has described. In doing so I shall confine myself to a few general observations, leaving the detailed appreciation, and perhaps, in some cases, criticisms, of the paper, to those who are far more able to deal with the technical aspect of it than I am.

Not only does the Complex Plan of experiments lead to increasing accuracy in experimental results, but also, as you will see, it has, in the hands of the Rothamsted workers, led to improvements in the theory of statistics. The history of those improvements has been well summarized by Mr. Yates. The essence of the new method is given on p. 189. The old idea was that we ought to arrange our experiments so that there was only one variable left. But, as Mr. Yates told us, Professor Fisher has shown that if we ask Nature a single question, she—*she*, observe—will often refuse to answer until some other topic has been discussed. Well, not only does factorial design enable us to estimate the effects of several variables at once, but also, according to Mr. Yates, it increases the accuracy. That is to me rather surprisingly brought out in the investigations on weighing on p. 210. There it is shown that if instead of carrying out a straightforward weighing in the old way, you weigh in groups, you can actually quadruple the accuracy and efficiency of the results. Then Mr. Yates shows how by a very simple nomenclature it is possible to apply algebraic methods to the results, and how the application of mathematics to statistics has led to great developments in recent years, such as the theory of variance and all the results which have followed therefrom. The discussion of these results I shall leave to others. I think the moral of the whole tale is that in all experimental work, whether agricultural or other, where we deal with large numbers, sound statistical advice is needed not only in the interpretation of the results but also in the planning of the experiments.

I should like to offer on behalf of us all a very cordial vote of thanks to Mr. Yates.