UNIVERSITY OF LONDON
GALTON LABORATORY, UNIVERSITY COLLEGE

EUGENICS LABORATORY MEMOIRS XLI

COMMENTARY ON R. A. FISHER'S PAPER ON

THE CORRELATION BETWEEN RELATIVES ON THE SUPPOSITION OF MENDELIAN INHERITANCE

[*Transactions of the Royal Society of Edinburgh* 52 (1918), 399–433]

BY

P. A. P. MORAN

Australian National University, Canberra

AND

C. A. B. SMITH

Galton Laboratory, University College London

PUBLISHED FOR THE
GALTON LABORATORY, UNIVERSITY COLLEGE LONDON
BY THE
CAMBRIDGE UNIVERSITY PRESS
BENTLEY HOUSE, 200 EUSTON ROAD, LONDON, N.W.1
AMERICAN BRANCH: 32 EAST 57TH STREET, NEW YORK, N.Y. 10022
1966

PUBLISHED BY
THE SYNDICS OF THE CAMBRIDGE UNIVERSITY PRESS
Bentley House, 200 Euston Road, London, N.W.1
American Branch: 32 East 57th Street, New York, N.Y. 10022
West African Office: P.M.B. 5181, Ibadan, Nigeria

Printed in Great Britain at the University Printing House, Cambridge
(Brooke Crutchley, University Printer)

PREFACE

The application of genetical principles to the study of human metrical characters, such as stature, was first attempted by Galton who, in 1887, used a method of correlation for measuring likeness between relatives. The theoretical basis of the results remained obscure until Mendelian principles of inheritance were applied. Karl Pearson's first attempt, in 1904, to account for the observed correlation values in this way was not satisfactory, but he succeeded in explaining the results in 1909 after the idea of random mating had been introduced into human genetics. It was not until 1918, however, that the matter was properly cleared up by Fisher's classical study, published in the *Proceedings of the Royal Society of Edinburgh*. Many aspects of the subject were dealt with in this paper, such as the effects of dominance and assortative mating on the correlation values. In some sections the exposition is very difficult to follow. The value of Fisher's contribution to the subject, however, is so great that Professor Moran and Professor Smith have thought it worth while to discuss his text in detail and criticize it where they think necessary. For this purpose the reprinting of the original paper is necessary and the running commentary provided should prove of great value both to students of genetics and of statistics.

L. S. PENROSE

INTRODUCTION

Sir Ronald Fisher's 1918 paper on the correlations between relatives is one of the classical papers of scientific literature. A few papers had previously appeared giving the expected values of the correlations on very simple Mendelian assumptions. Fisher succeeded in dealing with all the more obvious complications such as complete or partial recessivity, multiple allelism, epistacy, linkage, and assortative mating, and indeed with combinations of these, in one single paper. Since these complications are known or virtually certain to occur in real examples, this was a most important and necessary advance. Furthermore, this paper was published when Fisher was still only 28 years of age. The treatment suffers from a few minor defects. The model for assortative mating is rather a special one, though very ingenious; the argument dealing with linked genes is incomplete; and there is no mention of sex-linkage. But the first two of these defects are not easy to repair, and there has been no appreciable advance on Fisher's treatment of these points in the 47 years since his paper appeared.

It is also of interest that we can see in this paper the beginning of some of Fisher's most important statistical ideas. Thus he sets out the idea of partitioning variance into components. This presumably led to the Analysis of Variance. Fisher uses in this paper a technique which is very closely related to the analysis of variance applied to linear regression.

We are very much indebted to the Royal Society of Edinburgh and to Fisher's executor the Public Trustee of South Australia for permission to reproduce his original paper, and to Professor L. S. Penrose for his encouragement. The text of Fisher's paper has here been set in small type, enclosed in double quotations marks. (Some small changes have been made in the mathematical typography, in order to make it more consistent with the usual present-day practice used in the commentary. But there has been no alteration in the substance.) The commentary has been printed in larger type.

We hope that we have everywhere interpreted Fisher's ideas correctly and will succeed in making the paper more easy to follow.

P. A. P. MORAN
C. A. B. SMITH

The Correlation between relatives on the supposition of Mendelian Inheritance

By R. A. FISHER, B.A.

Communicated by Professor J. ARTHUR THOMSON

With Four Figures in Text

(*MS. received* 15 *June* 1918. *Read* 8 *July* 1918. *Issued separately* 1 *October* 1918)

CONTENTS

1. The superposition of factors distributed independently 6
2. Phase frequency in each array 10
3. Parental regression 10
4. Dominance deviations 10
5. Correlation for parent; genetic correlations 12
6. Fraternal correlation 15
7. Correlations for other relatives 17
8. Epistacy 22
9. Assortative mating 27
10. Frequency of phases 30
11. Association of factors 31
12. Conditions of equilibrium 32
13. Nature of association 34
14. Multiple allelomorphism 36
15. Homogamy and multiple allelomorphism 40
16. Coupling 44
17. Theories of marital correlation; ancestral correlations 50
18. Ancestral correlations (second and third theories) 51
19. Numerical values of association 51
20. Fraternal correlation 52
21. Numerical values for environment and dominance ratios; analysis of variance . 54
22. Other relatives 55
23. Numerical values (third theory) 56
24. Comparison of results 57
25. Interpretation of dominance ratio (diagrams) 58
26. Summary 61

" Several attempts have already been made to interpret the well-established results of biometry in accordance with the Mendelian scheme of inheritance. It is here attempted to ascertain the biometrical properties of a population of a more general type than has hitherto been examined, inheritance in which follows this scheme. It is hoped that in this way it will be possible to make a more exact analysis of the causes of human variability. The great body of available statistics show us that the deviations of a human measurement from its mean follow very closely the Normal Law of Errors, and, therefore, that the variability may be uniformly measured by the standard deviation corresponding to the square root of the mean square error. When there are two independent causes of variability capable of producing in an otherwise uniform population distributions with standard deviations σ_1 and σ_2, it is found that the distribution, when both causes act together, has a standard deviation $\sqrt{(\sigma_1^2+\sigma_2^2)}$."

This assumes that the causes act additively and not, for example, multiplicatively.

" It is therefore desirable in analysing the causes of variability to deal with the square of the standard deviation as the measure of variability. We shall term this quantity the Variance of the normal population to which it refers, and we may now ascribe to the constituent causes fractions or percentages of the total variance

which they together produce. It is desirable on the one hand that the elementary ideas at the basis of the calculus of correlations should be clearly understood, and easily expressed in ordinary language, and on the other that loose phrases about the 'percentage of causation', which obscure the essential distinction between the individual and the population, should be carefully avoided.

"Speaking always of normal populations, when the coefficient of correlation between father and son, in stature let us say, is r, it follows that for the group of sons of fathers of any given height the variance is a fraction, $1-r^2$, of the variance of sons in general. Thus if the correlation is 0·5, we have accounted by reference to the height of the father for one quarter of the variance of the sons."

This does not mean that one quarter of the variance is due to the direct genetic link between father and son. Some of the correlation may arise indirectly because of a resemblance between father and mother, and there is a direct genetic link between mother and son.

"For the remaining three quarters we must account by some other cause. If the two parents are independent, a second quarter may be ascribed to the mother. If father and mother, as usually happens, are positively correlated, a less amount must be added to obtain the joint contribution of the two parents, since some of the mother's contribution will in this case have been already included with the father's. In a similar way each of the ancestors makes an independent contribution, but the total amount of variance to be ascribed to the measurements of ancestors, including parents, cannot greatly exceed one half of the total. We may know this by considering the difference between brothers of the same fraternity: of these the whole ancestry is identical, so that we may expect them to resemble one another rather more than persons whose ancestry, identical in respect of height, consists of different persons. For stature the coefficient of correlation between brothers is about 0·54, which we may interpret* by saying that 54 per cent of their variance is accounted for by ancestry alone, and that 46 per cent must have some other explanation."

Fisher is using 'accounted' for in the technical sense that $R^2 = 0{\cdot}54$ is the multiple correlation of the measured value on the values of all ancestors. Fisher will show later that most of the remaining variability is also due to the parents, being caused by their heterozygosity. This does not contribute to the regression of child on parent, and thus, in the sense of the theory of regression, this part of the child's variability is not 'accounted for' by the parents' variability.

Suppose that x is a biological measurement on a son obtained by choosing a family at random out of a large population of families and choosing a son at random out of this family. Let x be measured from its mean and have variance σ^2. If X is the measurement on another son chosen from the same family the expected value of $(x-X)^2$ will be $2V$, where V is the variance of a son around the family mean. This mean value of $(x-X)^2$ is the mean over all families.

On the other hand, if x and X are the measurements on two brothers in the same family the mean value of $(x-X)^2$ taken over all families must be $2\sigma^2(1-r)$, where r is the correlation between brothers. Thus $2V = 2\sigma^2(1-r)$, and $V/\sigma^2 = 1-r$.

Suppose now that x and X are measurements on two parents, and z on their offspring. Then the proportion of the variance of z accounted for by the two parents is the multiple correlation of z

* The correlation is determined from the measurements of n individuals, $x_1, x_2, \ldots, x_n$, and of their brothers, $y_1, y_2, \ldots, y_r$; let us suppose that each pair of brothers is a random sample of two from an infinite fraternity, that is to say from all the sons which a pair of parents might conceivably have produced, and that the variance of each such fraternity is V, while that of the sons in general is σ. Then the mean value of $(x-y)^2$ will be $2V$, since each brother contributes the variance V. But expanding the expression, we find the mean value of both x^2 and y^2 is σ^2, while that of xy is $r\sigma^2$, where r is the fraternal correlation. Hence $2V = 2\sigma^2(1-r)$, or $V/\sigma^2 = 1-r$. Taking the values 0·5066 and 0·2804 for the parental and marital correlations, we find that the *heights of the parents alone* account for 40·10 per cent of the variance of the children, whereas the *total effect of ancestry*, deduced from the fraternal correlation, is 54·33 per cent. [All footnotes are from the original paper by Fisher.]

with x and X, i.e. in this case the correlation of z with $x+X$. The variances of x, X, and z are σ^2 each and the variance of $x+X$ is $2\sigma^2(1+r_m)$, where r_m is the correlation between x and X, i.e. the 'marital' correlation. The covariance of z with $(x+X)$ is the mean value of $z(x+X)$ which equals $2\sigma^2 r_p$ where r_p is the correlation between a son and a parent. The multiple correlation is therefore $r_p(1+r_m)^{-1}$ and in the particular case considered this is

$$0{\cdot}5066\,(1{\cdot}2804)^{-1} = 0{\cdot}3956,$$

which differs slightly from Fisher's value 0·4010. What Fisher calls the 'total effect of ancestry' is given by the observed fraternal correlation, which is 0·5433, because this is the square of the multiple correlation coefficient with all the ancestors and is therefore the fractional reduction in variance when all the ancestral values are held fixed. The standard errors of these estimates are not given.

" It is not sufficient to ascribe this last residue to the effects of environment. Numerous investigations by Galton and Pearson have shown that all measurable environment has much less effect on such measurements as stature. Further, the facts collected by Galton respecting identical twins show that in this case, where the essential nature is the same, the variance is far less. The simplest hypothesis, and the one which we shall examine, is that such features as stature are determined by a large number of Mendelian factors, and that the large variance among children of the same parents is due to the segregation of those factors in respect to which the parents are heterozygous. Upon this hypothesis we will attempt to determine how much more of the variance, in different measurable features, beyond that which is indicated by the fraternal correlation, is due to innate and heritable factors.

" In 1903 Karl Pearson devoted to a first examination of this hypothesis the twelfth of his Mathematical Contributions to the Theory of Evolution (' On a Generalised Theory of Alternative Inheritance, with special reference to Mendel's Laws,' *Phil. Trans.*, vol. CCIII, A, pp. 53–87. The subject had been previously opened by Udny Yule, *New Phytologist*, vol. I). For a population of n equally important Mendelian pairs, the dominant and recessive phases being present in equal numbers, and the different factors combining their effects by simple addition, he found that the correlation coefficients worked out uniformly too low. The parental correlations were $\frac{1}{2}$ and the fraternal $\frac{n}{1+n}$.*

" These low values, as was pointed out by Yule at the Conference on Genetics in 1906 (Horticultural Society's Report), could be satisfactorily explained as due to the assumption of complete dominance. It is true that dominance is a very general Mendelian phenomenon, but it is purely somatic, and if better agreements can be obtained without assuming it in an extreme and rigorous sense, we are justified in testing a wider hypothesis. Yule, although dealing with by no means the most general case, obtained results which are formally almost general. He shows the similarity of the effects of dominance and of environment in reducing the correlations between relatives, but states that they are identical, an assertion to which, as I shall show, there is a remarkable exception, which enables us, as far as existing statistics allow, to separate them and to estimate how much of the total variance is due to dominance and how much to arbitrary outside causes.

" In the following investigation we find it unnecessary to assume that the different Mendelian factors are of equal importance, and we allow the different phases of each to occur in any proportions consistent with the

* The case of the fraternal correlations has been unfortunately complicated by the belief that the correlation on a Mendelian hypothesis would depend on the number of the fraternity. In a family, for instance, in which four Mendelian types are liable to occur in equal numbers, it was assumed that of a family of four, one would be of each type; in a family of eight, two of each type; and so on. If this were the case, then in such families, one being of the type A would make it less likely, in small families impossible, for a second to be of this type. If, as was Mendel's hypothesis, the different qualities were carried by different gametes, each brother would have an independent and equal chance of each of the four possibilities. Thus the formulae giving the fraternal correlations in terms of the number of the fraternity give values too small. The right value on Mendel's theory is that for an infinite fraternity. As Pearson suggested in the same paper, ' probably the most correct way of looking at any fraternal correlation table would be to suppose it a random sample of all pairs of brothers which would be obtained by giving a large, or even indefinitely large, fertility to each pair, for what we actually do is to take families of varying size and take as many pairs of brothers as they provide.' In spite of this, the same confusing supposition appears in a paper by Snow ' On the Determination of the Chief Correlations between Collaterals in the Case of a Simple Mendelian Population Mating at Random' (E. C. Snow, B.A., *Proc. Roy. Soc.* June 1910); and in one by John Brownlee, ' The Significance of the Correlation Coefficient when applied to Mendelian Distributions' (*Proc. Roy. Soc. Edinb.* Jan. 1910).

conditions of mating. The heterozygote is from the first assumed to have any value between those of the dominant and the recessive, or even outside this range, which terms therefore lose their polarity, and become merely the means of distinguishing one pure phase from the other. In order to proceed from the simple to the complex we assume at first random mating, the independence of the different factors, and that the factors are sufficiently numerous to allow us to neglect certain small quantities."

Although Fisher states that random mating is assumed at first, the theory is developed in terms more general than this and he is careful to state when the additional assumption is introduced. He also assumes for the present that each measured character is the result of summing a large number of small factors which are independent, i.e. that there is no linkage. It then follows from the standard properties of means, variances and covariances that the mean value of the character in the population is equal to the sum of the means of the individual small factors, the variance is similarly the sum of the individual variances, and the same is true of the covariances.

Suppose that for the particular factor considered the two possible alleles are A_1 and A_2. We then have the following table:

Zygote	A_1A_1	A_1A_2	A_2A_2
Phenotypic effect	a	d	$-a$
Frequency	P	$2Q$	R

If the individuals concerned had been produced by a process involving random mating and no selection we would have

$$PR = Q^2,$$

and

$$p = P+Q, \quad q = Q+R,$$

would be the gene frequencies of the A_1 and A_2 genes so that

$$P = p^2, \quad Q = pq, \quad R = q^2.$$

As assortative mating is considered later, it is more convenient to develop the theory in terms of P, Q and R without assuming the Hardy–Weinberg formula except when random mating is explicitly asserted.

The variance α^2 given by (I) is the contribution of this factor to the total variance σ^2 whether or not the distribution is normal. The fact that the distribution of the sum of all factors will be approximately normally distributed (particularly if measured after a suitable transformation) will follow from the version of the Central Limit Theorem which proves asymptotic normality for a sum of independent random variables each of which is 'individually negligible' in a certain precise sense. The calculation of the third and fourth moment here is merely illustrative.

" 1. Let us suppose that the difference caused by a single Mendelian factor is represented in its three phases by the difference of the quantities a, d, $-a$, and that these phases exist in any population with relative frequency P, $2Q$, R, where $P+2Q+R = 1$.

" Then a population in which this factor is the only cause of variability has its mean at

$$m = Pa+2Qd-Ra,$$

so that

$$P(a-m)+2Q(d-m)-R(a+m) = 0.$$

Let now

$$P(a-m)^2+2Q(d-m)^2+R(a+m)^2 = \alpha^2 \qquad \text{(I)}$$

α^2 then is the variance due to this factor, for it is easily seen that when two such factors are combined at random, the mean square deviation from the new mean is equal to the sum of the values of α^2 for the two factors

separately. In general the mean square deviation due to a number of such factors associated at random will be written

$$\sigma^2 = \Sigma\alpha^2. \tag{II}$$

" To justify our statement that α^2 is the contribution which a single factor makes to the total variance, it is only necessary to show that when the number of such factors is large the distributions will take the normal form.

" If now we write

$$\mu_3 = P(a-m)^3 + 2Q(d-m)^3 - R(a+m)^3,$$

$$\mu_4 = P(a-m)^4 + 2Q(d-m)^4 + R(a+m)^4,$$

and if M_3 and M_4 are the third and fourth moments of the population, the variance of which is due solely to the random combination of such factors, it is easy to see that

$$M_3 = \Sigma\mu_3,$$

$$M_4 - 3\sigma^4 = \Sigma(\mu_4 - 3\alpha^4).$$

Now the departure from normality of the population may be measured by means of the two ratios

$$\beta_1 = \frac{M_3^2}{\sigma^6} \quad \text{and} \quad \beta_2 = \frac{M_4}{\sigma^4}.$$

The first of these is

$$(\Sigma\mu_3)^2/(\Sigma\alpha^2)^3,$$

and is of the order $1/n$, where n is the number of factors concerned, while the second differs from its Gaussian value 3 also by a quantity of the order $1/n$."

In sections 2 and 3 the following problem is considered. Suppose that the measurement x (measured from the population mean) is the sum of the effects of a large number of independently segregating factors. For a parent (say a father) and an offspring (say a son) these measurements will be distributed, to a high degree of approximation, in a bivariate normal distribution, and we wish to calculate the regression coefficient of the value for the son on the value for the father. This is done by an ingenious approximate argument. In this it is assumed that the parents mate at random but not necessarily the grandparents.

x (whose variance is σ^2) is the sum of a large number of independent factor pairs of which a typical one is (A_1, A_2) whose contribution to the variance is α^2. Suppose the proportions of (A_1A_1) (A_1A_2) and (A_2A_2) in the whole population are $\bar{P}$, $\bar{Q}$, $\bar{R}$. We now choose a particular value x for the father. In the subpopulation of fathers having this value, the frequencies of (A_1A_1) (A_1A_2) and (A_2A_2) will be different and we write P, Q and R for them. Our first task is to calculate these.

To do this, consider a population of fathers in which all the factors have frequencies the same as in the above population except for the one factor considered for which all individuals are to be heterozygotes (A_1A_2). The variance of this population is then $\sigma^2 - \alpha^2$ since the component variance, α^2, due to A_1, A_2, has been removed in this way. If we now modify this in the manner described and use the fact that the distribution of x is normal we see that the frequencies of (A_1A_1), (A_1A_2) and (A_2A_2) must be

$$\bar{P}\exp\left\{-\frac{(x-a+m)^2}{2(\sigma^2-\alpha^2)}\right\}, \quad \text{etc.},$$

in order to get the previously considered population. From this we obtain (III) which is an approximation obtained by supposing that

$$\alpha^2/\sigma^2 \quad \text{and} \quad x^2/\sigma^2$$

are small. These give the proportions of the three types in a population of fathers with value x.

We suppose that these fathers mate at random with the general population. For the particular factor considered we then get Table A, in which each cell gives the values of the possible offspring with their frequencies.

TABLE A

Mother, from rest of population			Father array: A_1A_1, a, P	Father array: A_1A_2, d, $2Q$	Father array: A_2A_2, $-a$, R
A_1A_1	a	$\bar{P}$	a $P\bar{P}$	a $\bar{P}Q$; d $\bar{P}Q$	d $\bar{P}R$
A_1A_2	d	$2\bar{Q}$	a $P\bar{Q}$; d $P\bar{Q}$	a $Q\bar{Q}$; d $2Q\bar{Q}$; $-a$ $Q\bar{Q}$	d $\bar{Q}R$; $-a$ $\bar{Q}R$
A_2A_2	$-a$	$\bar{R}$	d $P\bar{R}$	d $Q\bar{R}$; $-a$ $Q\bar{R}$	$-a$ $R\bar{R}$

The sons therefore have values a, d and $-a$ with probabilities

$$\bar{P}P+\ \bar{P}Q+\ P\bar{Q}+\ Q\bar{Q},$$

$$\bar{P}Q\ +\ P\bar{Q}\ +\ P\bar{R}\ +\ \bar{P}R\ +\ \bar{R}Q+R\bar{Q}+2Q\bar{Q},$$

$$Q\bar{Q}+\ \bar{Q}R+\ Q\bar{R}+\ R\bar{R}.$$

We now insert the values given by (III) and ignore terms of higher order than the first in x/σ^2 and we obtain the formulae given at the beginning of paragraph 3. Multiplying these by a, d and $-a$ and adding, we find that the expected value of the mean of the offspring is

$$2d(\bar{P}\bar{R}-\bar{Q}^2)+\frac{x}{\sigma^2}[\bar{P}\bar{Q}(a-d)^2+2\bar{P}\bar{R}(a^2-d^2)+\bar{Q}\bar{R}(a+d)^2+2(\bar{P}\bar{R}-\bar{Q}^2)\,d(d-m)]. \qquad \text{(III}a\text{)}$$

(A factor 2 multiplying $(\bar{P}\bar{R}-\bar{Q}^2)$ inside the square bracket is omitted in Fisher.) Note that in order to obtain (IIIa) it is necessary to use the result

$$m(P+2Q+R)=aP+2dQ-aR.$$

If the *parents* are the result of a mating at random

$$\bar{P}\bar{R}-\bar{Q}^2=0,$$

and (IIIa) simplifies to (IV).

Thus (IV) has been obtained by an approximate argument. However (IV) is exact in the sense that it gives the ratio of the part of the covariance, which is due to the factor considered, to σ^2. This means that if x is the value of the father and X of the son, where

$$x=x_1+x_2+\ldots, \quad X=X_1+X_2+\ldots,$$

and x_i, X_i are the values of the contribution made by factor i, then

$$\text{covariance}\ (x_i, X_i)=[\bar{P}\bar{Q}(a-d)^2+2\bar{P}\bar{R}(a^2-d^2)+\bar{Q}\bar{R}(a+d)^2].$$

(IIIa) is exact whether $\bar{P}\bar{R}-\bar{Q}^2=0$, or not.

We shall now prove this, and in doing so we shall revert to the notation P, Q, R instead of $\bar{P}$, $\bar{Q}$, $\bar{R}$ as Fisher does this in paragraph 4 onwards. In this way we will see that (IIIa) is also

exact. In the association table (Table B), the columns correspond to the genotypes of the male parent and the rows to the genotype of the female parent. In each cell the genotypes of the offspring with their probabilities are given, assuming random mating in the parents. From this we can immediately extract an association table for parent/offspring (Table C).

TABLE B

Female parent			Male parent A_1A_1 a P	Male parent A_1A_2 d $2Q$	Male parent A_2A_2 $-a$ R
A_1A_1	a	P	a P^2	a d PQ PQ	d PR
A_1A_2	d	$2Q$	a d PQ PQ	a d $-a$ Q^2 $2Q^2$ Q^2	d $-a$ QR QR
A_2A_2	$-a$	R	d PR	d $-a$ QR QR	$-a$ R^2

TABLE C

Offspring	Parent A_1A_1	Parent A_1A_2	Parent A_2A_2
A_1A_1	P^2+PQ	$PQ+Q^2$	0
A_1A_2	$PQ+PR$	$PQ+2Q^2+QR$	$PR+QR$
A_2A_2	0	Q^2+QR	$QR+R^2$

Notice that unlike Tables A and B, Table C is not symmetric about the leading diagonal but does still have another type of symmetry about the other diagonal resulting from the symmetric role of the two factors, A_1 and A_2 (the two previous tables of course also have this type of symmetry).

From Table C in turn we can find the mean value of the offspring multiplied by the probability of the parent, for each of the parental types, and this is shown in Table D. The sum of the third column gives the mean value of the offspring which is

$$2d(PR-Q^2).$$

TABLE D

Parental type and its value		Probability of parental type multiplied by mean value of offspring
A_1A_1	a	$aP^2+aPQ+dPQ+dPR$
A_1A_2	d	$aPQ-aQR+d(PQ+2Q^2+QR)$
A_2A_2	$-a$	$dPR+dQR-aQR-aR^2$

The covariance uncorrected for the means is the sum of the products of the first and second columns. Calculating this and subtracting the correction for the means which is

$$m\{m+2d(PR-Q^2)\},$$

we verify the formula before (IV) which is therefore exact when corrected as in (IIIa).

When the parents have been produced by random mating we have $PR-Q^2=0$ which is the Hardy–Weinberg relation and we can then write

$$P=p^2,\quad Q=pq,\quad R=q^2,$$

where $p=P+Q$, $q=Q+R$.

" 2. If there are a great number of different factors, so that σ is large compared to every separate α, we may investigate the proportions in which the different phases occur in a selected array of individuals. Since the deviation of an individual is simply due to a random combination of the deviations of separate factors, we must expect a given array of deviation, let us say x, to contain the phases of each factor in rather different proportions to those in which they exist in the whole population. The latter will be represented now by $\bar{P}$, $2\bar{Q}$, $\bar{R}$, while P, $2Q$, R stand for the proportions in some particular array under consideration.

" Consider a population which is the same in every respect as the one we are dealing with save that all its members have one particular factor in the heterozygous phase, and let us modify it by choosing of each array a proportion $\bar{P}$ which are to become dominants and to increase by $a-d$, and a proportion $\bar{R}$ which become recessive and diminish by $a+d$: the mean is thereby moved to the extent $m-d$.

" Of those which after this modification find themselves in the array with deviation x, the dominants formerly had a deviation $x-a+m$, the heterozygotes $x-d+m$, and the recessives $x+a+m$, and since the variance of the original population was $\sigma^2-\alpha^2$, the frequencies of these three types are in the ratio

$$\bar{P}\exp\left\{-\frac{(x-a+m)^2}{2(\sigma^2-\alpha^2)}\right\}:2\bar{Q}\exp\left\{-\frac{(x-d+m)^2}{2(\sigma^2-\alpha^2)}\right\}:\bar{R}\exp\left\{-\frac{(x+a+m)^2}{2(\sigma^2-\alpha^2)}\right\},$$

or, when σ is great compared to α, so that α^2/σ^2 may be neglected,

$$\left.\begin{aligned}P&=\bar{P}\left[1+\frac{x}{\sigma^2}(a-m)\right]\\ Q&=\bar{Q}\left[1+\frac{x}{\sigma^2}(d-m)\right]\\ R&=\bar{R}\left[1-\frac{x}{\sigma^2}(a+m)\right]\end{aligned}\right\}\qquad\text{(III)}$$

giving the proportions in which the phases occur in the array of deviation x.

" 3. Hence the members of this array mating at random will have offspring distributed in the three phases in the proportion

$$\bar{P}^2\left[1+\frac{x}{\sigma^2}(a-m)\right]+\bar{P}\bar{Q}\left[2+\frac{x}{\sigma^2}(a-m+d-m)\right]+\bar{Q}^2\left[1+\frac{x}{\sigma^2}(d-m)\right],$$

$$\bar{P}\bar{Q}\left[2+\frac{x}{\sigma^2}(a-m+d-m)\right]+2\bar{Q}^2\left[1+\frac{x}{\sigma^2}(d-m)\right]+\bar{P}\bar{R}\left[2-\frac{x}{\sigma^2}(2m)\right]+\bar{Q}\bar{R}\left[2+\frac{x}{\sigma^2}(d-m-a-m)\right],$$

$$\bar{Q}^2\left[1+\frac{x}{\sigma^2}(d-m)\right]+\bar{Q}\bar{R}\left[2+\frac{x}{\sigma^2}(d-m-a-m)\right]+\bar{R}^2\left[1-\frac{x}{\sigma^2}(a+m)\right],$$

and therefore the deviation of the mean of the offspring is

$$2d(\bar{P}\bar{R}-\bar{Q}^2)+\frac{x}{\sigma^2}[\bar{P}\bar{Q}(a-d)^2+2\bar{P}\bar{R}(a^2-d^2)+\bar{Q}\bar{R}(a+d)^2+(\bar{P}\bar{R}-\bar{Q}^2)d(d-m)].$$

" Omitting the terms in $(\bar{P}\bar{R}-\bar{Q}^2)$, which for random mating is zero, the regression due to a single factor is

$$\frac{x}{\sigma^2}[\bar{P}\bar{Q}(\alpha-d)^2+2\bar{P}\bar{R}(a^2-d^2)+\bar{Q}\bar{R}(a+d)^2].\qquad\text{(IV)}$$

" 4. To interpret this expression, consider what is involved in taking a, d, $-a$ as representing the three phases of a factor. Genetically the heterozygote is intermediate between the dominant and the recessive, somatically it differs from their mean by d. The steps from recessive to heterozygote and from heterozygote to dominant are genetically identical, and may change from one to the other in passing from father to son. Somatically the steps are of different importance, and the soma to some extent disguises the true genetic nature. There is in dominance a certain latency. We may say that the somatic effects of identical genetic changes are not additive, and for this reason the genetic similarity of relations is partly obscured in the statistical aggregate. A similar deviation from the addition of superimposed effects may occur between

different Mendelian factors. We may use the term Epistacy to describe such deviation, which although potentially more complicated, has similar statistical effects to dominance. If the two sexes are considered as Mendelian alternatives, the fact that other Mendelian factors affect them to different extents may be regarded as an example of epistacy."

The value, d, for the heterozygote A_1A_2 will not be exactly intermediate between the value a for A_1A_1 and the value $-a$ for A_2A_2 unless $d = 0$. Fisher proposes to replace these by values for which the heterozygote is $c+b, c, c-b$. These values are fitted by least squares, i.e. by minimizing the sum (Fisher uses S without a suffix for summation),

$$S_1 = P(c+b-a)^2 + 2Q(c-d)^2 + R(c-b+a)^2.$$

This procedure is equivalent to considering the linear regression of the measured value on the number of A_2 genes present, and the reason for its usefulness will appear later. To minimize S_1 we have to solve the equations

$$\frac{1}{2}\frac{\partial S_1}{\partial b} = P(c+b-a) - R(c-b+a) = 0,$$

$$\frac{1}{2}\frac{\partial S_1}{\partial c} = P(c+b-a) + 2Q(c-d) + R(c-b+a) = 0.$$

The solution is

$$c = \frac{(P+R)Qd}{T}, \quad b = a - \frac{Q(P-R)d}{T},$$

where $T = PQ + 2PR + QR$.

Fisher's formula for b should have the first plus sign changed to minus. Notice that b and c depend not only on a and d, but also on the frequencies P, $2Q$, R.

Notice also that if $PR = Q^2$, $T = Q$.

Using these values we find the deviations from the regression line for A_1A_1, A_1A_2, and A_2A_2 to be

$$c+b-a = 2RQd/T, \quad c-d = -2PRd/T, \quad c-b+a = 2PQd/T. \tag{IVa}$$

These deviations have the expected value

$$P(c+b-a) + 2Q(c-d) + R(c-b+a) = 2dT^{-1}(PRQ - 2QPR + RPQ) = 0.$$

Their variance is therefore

$$\begin{aligned}\delta^2 &= P(c+b-a)^2 + 2Q(c-d)^2 + R(c-b+a)^2 \\ &= 4PQRd^2/T.\end{aligned} \tag{IVb}$$

This is also by definition the minimum value of S_1, as follows from the ordinary least squares regression theory.

The covariance between the 'representative values' and the 'deviations from linearity' is then the mean product (since the mean deviation is zero). This is

$$\begin{aligned}&P(c+b-a)(c+b) + 2Q(c-d)c + R(c-b+a)(c-b) \\ &= 2dT^{-1}\{[PQR(c+b) - 2QPRc + RPQ(c-b)] \\ &= 0.\end{aligned}$$

Thus the correlation is zero as again follows from the usual regression theory.

The total 'genotypic' variance is

$$\alpha^2 = P(a-m)^2 + 2Q(d-m)^2 + R(-a-m)^2,$$

and can be decomposed into two parts. The first of these is the variance of the representative values

$$\beta^2 = P(c+b-m)^2 + 2Q(c-m)^2 + R(c-b-m)^2,$$

which is nowadays called the 'genetic' variance due to the A_1, A_2 genes. The second is the variance of the 'dominance deviations', δ^2, as given by (IV b) above. We can verify algebraically that

$$\alpha^2 = \beta^2 + \delta^2,$$

which is again a consequence of the usual regression theory, especially when the latter is presented in an analysis of variance table.

If random mating holds, $T = Q$, and $PR = Q^2$ so that

$$\alpha^2 = 2a^2Q - 4Q(P-R)\,ad + 2Q(P+R)\,d^2$$

and

$$\beta^2 = 2a^2Q - 4Q(P-R)\,ad + 2Q(P-R)^2\,d^2.$$

(Fisher has $2a^2Q^2$ in this formula (formula (VI)) which is wrong.) Then

$$\alpha^2 - \beta^2 = 4Q^2d^2 = \delta^2.$$

The total variance, σ^2, of the character in the population is the sum, $\Sigma\alpha^2$, over all pairs of genes like A_1, A_2, since we suppose the character is additive. Fisher writes

$$\tau^2 = \Sigma\beta^2, \quad \epsilon^2 = \Sigma\delta^2,$$

so that $\sigma^2 = \tau^2 + \epsilon^2$.

" The contributions of imperfectly additive genetic factors divide themselves for statistical purposes into two parts: an additive part which reflects the genetic nature without distortion, and gives rise to the correlations which one obtains; and a residue which acts in much the same way as an arbitrary error introduced into the measurements. Thus, if for a, d, $-a$ we substitute the linear series

$$c+b, c, c-b,$$

and choose b and c in such a way that

$$P(c+b-a)^2 + 2Q(c-d)^2 + R(c-b+a)^2$$

is a minimum, we find for this minimum value δ^2,

$$\delta^2 = \frac{4PQR\,d^2}{PQ+2PR+QR},$$

which is the contribution to the variance of the irregular behaviour of the soma; and for the contribution of the additive part, β^2, where

$$\beta^2 = P(c+b-m)^2 + 2Q(c-m)^2 + R(c-b-m)^2,$$

we obtain

$$\beta^2 = 2b^2(PQ+2PR+QR),$$

and since

$$b = a + \frac{Q(P-R)\,d}{PQ+2PR+QR},$$

we have

$$\beta^2 = 2a^2(PQ+2PR+QR) - 4Q(P-R)\,ad + \frac{2Q^2(P-R)^2\,d^2}{PQ+2PR+QR}.$$

" 5. These expressions may be much simplified by using the equation

$$Q^2 = PR,$$

for then

$$\delta^2 = 4Q^2d^2 \tag{V}$$

$$\beta^2 = 2a^2Q^2 - 4Q(P-R)\,ad + 2Q(P-R)^2d^2, \tag{VI}$$

which appears in the regression in Article 3 (IV), and

$$\alpha^2 = 2a^2Q(P-R)\,ad + 2Q(P+R)\,d^2. \tag{VII}$$

"In general $\alpha^2 = \beta^2 + \delta^2$,

and if $\sigma^2 = \Sigma\alpha^2$, (VIII)

$\tau^2 = \Sigma\beta^2$, (IX)

and $\epsilon^2 = \Sigma\delta^2$, (X)

then $\sigma^2 = \tau^2 + \epsilon^2$."

The reasons for introducing this type of regression analysis are most easily seen from later formulations of the problem by Malécot (*Les mathématiques de l'hérédité*, Paris (1948)), Li and Sacks (*Biometrics*, **10**, (1954), 347–360). At any given locus any individual has two genes which can be distinguished by their origin one from the individual's father, and one from his mother. The effect produced by this pair of genes can be split up into three components in the following way:

First gene	Second gene	Frequency	$x = x_1 + x_2 + x_3$
A_1	A_1	P	$a = \frac{1}{2}(c+b) + \frac{1}{2}(c+b) + (a-c-b)$
A_1	A_2	Q	$d = \frac{1}{2}(c+b) + \frac{1}{2}(c-b) + (d-c)$
A_2	A_1	Q	$d = \frac{1}{2}(c-b) + \frac{1}{2}(c+b) + (d-c)$
A_2	A_2	R	$-a = \frac{1}{2}(c-b) + \frac{1}{2}(c-b) + (-a-c+b)$

The first component is $\frac{1}{2}(c+b)$ or $\frac{1}{2}(c-b)$ according as the first gene is A_1 or A_2, and similarly for the second gene. The third component is a deviation from linearity. Fisher's 'representative value' is $x_1 + x_2$. With random mating we find from what has gone previously, that

$$\tfrac{1}{2}\beta^2 = \text{var}\,(x_1) = \text{var}\,(x_2), \quad \delta^2 = \text{var}\,(x_3),$$

and the three covariances between the x's are zero. The correlation between $(x_1+x_2+x_3)$ and (x_1+x_2) is then

$$\frac{\text{var}\,(x_1+x_2)}{\{\text{var}\,(x_1+x_2+x_3)\,\text{var}\,(x_1+x_2)\}^{\frac{1}{2}}} = \frac{\beta^2}{(\alpha^2\beta^2)^{\frac{1}{2}}} = \frac{\beta}{\alpha}.$$

Now consider a parent and offspring with values $(x_1+x_2+x_3)$ and $(X_1+X_2+X_3)$ respectively. We make the convention that the 'first' gene (which results in the contributions x_1 and X_1) is the gene which this parent hands on to the offspring, so that $x_1 = X_1$. The second genes in the two individuals are A_1 and A_2, with probabilities p and q, independently of each other. Thus $x_1 = X_1$, x_2, and X_2 are distributed independently of each other, and so are the pairs (x_2, X_3) and (x_3, X_2). $x_1 = X_1$ is uncorrelated with x_3 and X_3 from what has been proved above.

It can also be shown that x_3 and X_3 are uncorrelated. This can be done as follows. Suppose that the gene passed from parent to offspring is A_1. Then using the above table and the fact that $P = p^2$, $Q = pq$,

$$E(x_3 \mid \text{first gene is } A_1) = p(a-c-b) + q(d-c)$$
$$= p(2RQdT^{-1}) + q(-2PRdT^{-1}) = 0.$$

Similarly, $$E(X_3 \mid \text{first gene is } A_1) = 0.$$

Then
$$E(x_3 X_3 \mid \text{first gene is } A_1)$$
$$= E(x_3 \mid \text{first gene is } A_1)\, E(X_3 \mid \text{first gene is } A_1)$$
$$= 0.$$

The same holds if the first gene is A_2 and thus

$$E(x_3) = E(X_3) = E(x_3 X_3),$$

so that $$\text{cov}\,(x_3, X_3) = 0.$$

The five variates $x_1 = X_1, x_2, x_3, X_2, X_3$ are thus uncorrelated in pairs, so that the correlation between x and X arises only through the pair x_1, X_1. Then

$$\text{cov}(x, X) = \text{cov}(x_1, X_1) = \text{var}(x_1) = \tfrac{1}{2}\beta^2.$$

Fisher does not consider the components x_1 and X_1 separately but the 'representative values' (x_1+x_2) and (X_1+X_2). Thus from his point of view the correlation between parent and offspring arises solely from that of the representative values.

Most pairs of relatives in a population can share a gene which may be passed directly from one relative to another as with father and son, or which may come from a common ancestor as with brothers. They are then said to have genes which are 'identical by descent', as distinct from pairs of genes which may be identical by chance. We can say that father and son have 'one gene in common'. Similarly, uncle and nephew have probability $\frac{1}{2}$ of having a gene in common, first cousins have probability $\frac{1}{4}$ of having a gene in common, and so on. Then arguing as above the correlation between $x_1, x_2, x_3, X_1, X_2, X_3$ (where $x_1+x_2+x_3$ and $X_1+X_2+X_3$ refer to the two individuals) are all zero except that $\text{cov}(x_1, X_1) = \frac{1}{2}u\beta^2$,

where u is the probability of sharing a gene.

With pairs of sibs, or double first cousins, the situation is more complicated, since the individual can then share two genes at once. In such a case each x_r may be correlated with X_r $(r = 1, 2, 3)$, but if $r \neq s$, the pairs $(x_r, x_s), (X_r, X_s), (x_r, X_s)$ are uncorrelated. Thus Fisher remarks that with sibs and other such cases, it is necessary to take into account the correlation between the 'dominance deviations' x_3 and X_3.

A valuable general theory of this approach is given by Trustrum (*Proc. Camb. Phil. Soc.* **57** (1961), 315–320).

" The regression due to a single factor of the mean of the offspring of parents of a given array is

$$\frac{x^2}{\sigma^2} \cdot \frac{\beta^2}{2},$$

and adding up the effects of all factors we find $\dfrac{x}{\sigma^2} \cdot \dfrac{\tau^2}{2}$,

so that the parental correlation for a static population mating at random is simply

$$\frac{1}{2} \cdot \frac{\tau^2}{\sigma^2}. \qquad \text{(XI)}$$

We may regard this formula otherwise. The correlation between the actual somatic measurements such as $a, d, -a$, and the representative linear quantities $c+b, c, c-b$ is τ/σ. Thus the correlation of parent and child is made up of three factors, two of them representing the relations between the real and the representative measurements, and the third the correlation between the representative measurements of the two relatives. Thus the effect of dominance is simply to reduce certain relationship correlations in the ratio τ^2/σ^2.

" The values of the correlations between the representative measurements for random mating, which may be called the genetic correlations, are given in the accompanying table:

Generations	Half 2nd cousin	Half 1st cousin	Half brother	Ancestral line	Brother	1st cousin	2nd cousin
Own	1/64	1/16	1/4	1	1/2	1/8	1/32
Father's	1/128	1/32	1/8	1/2	1/4	1/16	1/64
Grandfather's	1/256	1/64	1/16	1/4	1/8	1/32	1/128
Great-grandfather's	1/512	1/128	1/32	1/8	1/16	1/64	1/256
Great-great-grandfather's	1/1024	1/256	1/64	1/16	1/32	1/128	1/512

" 6. The above reasoning as to the effects of dominance applies without modification to the ancestral line, but in a special class of collaterals requires reconsideration. The reason is that the deviations from linearity are now themselves correlated. In other words, a father who is heterozygote instead of recessive may have offspring who show a similar variation; but they may also be changed from heterozygote to dominant. In the case of siblings, however, whichever change takes place in one is more likely to occur in the other.

" Thus, writing i, j, k for the deviations

$$a-m, \quad d-m, \quad -(a+m),$$

so that

$$iP+2jQ+kR=0 \qquad \text{(XII)}$$

and p^2, pq, q^2 for P, Q, R, we can draw up association tables for different pairs of relatives, and readily obtain the correlations between them by substituting the fractions in the nine sections of the table as coefficients of a quadratic function in i, j, k.

" Thus the association table between parent and child is

p^3	p^2q	—
p^2q	$pq(p+q)$	pq^2
—	pq^2	q^3

from which we obtain the quadratic

$$p^3i^2+2p^2qij+pq(p+q)j^2+2pq^2jk+q^3k^2,$$

which is equal to

$$\frac{1}{4pq}(p^2i-q^2k)^3=\tfrac{1}{2}\beta^2,$$

The association table for parent and child given by Fisher above has its columns corresponding to the three genotypes A_1A_1, A_1A_2, A_2A_2 respectively, and hence to the deviations i, j, k. The rows have a similar meaning for the offspring. The entries in the table are the respective probabilities of occurrence of all combinations of father and offspring; e.g. the combination father A_1A_1, offspring A_1A_1, has probability p^3. The entries can be found by putting $P=p^2$, $Q=pq$, $R=q^2$ in Table C, using $p+q=1$. The 'quadratic' under the table is the covariance (a word which he had presumably not yet invented). This can be found directly from its definition as a mean product of deviations, i.e. as

Σ (prob) (parent's deviation) (offspring's deviation)

$$=p^3.i.i.+p^2q.j.i+\ldots$$
$$=p^3i^2+2p^2qij+pq(p+q)j^2+2pq^2jk+q^3k^2$$

(there being a misprint in Fisher's text). On substituting for j, using (XII), this becomes

$$\frac{1}{4pq}(ip^2-kq^2)^2=\tfrac{1}{2}\beta^2,$$

the bracketed expression being squared and not cubed as in Fisher's text.

To obtain the variance of father and offspring we use the formula for β^2 given before. Then

$$\beta^2=2a^2Q-4Q(P-R)\,ad+2Q(P-R)^2d^2$$
$$=2a^2pq-4pq(p-q)\,ad+2pq(p-q)^2d^2$$
$$=2pq\{a-(p-q)\,d\}^2.$$

Now

$$p^2i+2pqj+q^2k=0,$$

and since

$$a-(p-q)\,d=\tfrac{1}{2}(i-k)-(p-q)\{j-\tfrac{1}{2}(i+k)\},$$

we find

$$2\{a-(p-q)\,d\}=i-k+\frac{p-q}{pq}(p^2i+q^2k)+(p-q)(i+k)$$
$$=\frac{1}{pq}(ip^2-kq^2).$$

Thus the variance is

$$\beta^2 = \frac{1}{2pq}\{ip^2 - kq^2\}^2,$$

which is twice the covariance obtained above.

" while for brother and brother we have the table

$p^2(p+\frac{1}{2}q)^2$	$p^2q(p+\frac{1}{2}q)$	$\frac{1}{4}p^2q^2$
$p^2q(p+\frac{1}{2}q)$	$pq(p^2+3pq+q^2)$	$pq^2(\frac{1}{2}p+q)$
$\frac{1}{4}p^2q^2$	$pq^2(\frac{1}{2}p+q)$	$q^2(\frac{1}{2}p+q)^2$

which gives us a quadratic expression exceeding that for the parental correlation by the terms

$$\frac{p^2q^2}{4}(i^2 - 2ij + 4j^2 + 2ik - 2jk + k^2),$$

which are equal to $\frac{1}{4}\delta^2$, and therefore give for the fraternal correlation

$$\frac{1}{2\sigma^2}(\tau^2 + \tfrac{1}{2}\epsilon^2)."$$

To obtain the brother–brother table we consider the table given before (Table B) of all possible offspring of two randomly mated parents, and examine all possible fraternities. Then an (A_1A_1, A_1A_1) fraternity can arise out of a crossing $A_1A_1 \times A_1A_1$ with probability P^2, or out of a crossing $A_1A_2 \times A_1A_1$ with probability $\frac{1}{4}(2PQ + 2PQ) = PQ$, or finally out of a crossing $A_1A_2 \times A_1A_2$ with probability $\frac{1}{16}(4Q^2) = \frac{1}{4}Q^2$. This gives us the cell in the first row and first column of Table E and the others are obtained similarly.

TABLE E

	Brother		
Brother	i	j	k
i	$P^2+PQ+\frac{1}{4}Q^2$	$PQ+\frac{1}{2}Q^2$	$\frac{1}{4}Q^2$
j	$PQ+\frac{1}{2}Q^2$	$PQ+2PR+Q^2+QR$	$\frac{1}{2}Q^2+QR$
k	$\frac{1}{4}Q^2$	$\frac{1}{2}Q^2+QR$	$\frac{1}{4}Q^2+QR+R^2$

Notice this is symmetric about the leading diagonal and symmetric about the other diagonal on interchanging P and R. On substituting for P, Q and R we get Fisher's table. To save algebraic labour we subtract the previous table and get an array of the form:

$\frac{1}{4}p^2q^2$	$-\frac{1}{2}p^2q^2$	$\frac{1}{4}p^2q^2$
$-\frac{1}{2}p^2q^2$	p^2q^2	$-\frac{1}{2}p^2q^2$
$\frac{1}{4}p^2q^2$	$-\frac{1}{2}p^2q^2$	$\frac{1}{4}p^2q^2$

from which we immediately get the expression

$$\tfrac{1}{4}p^2q^2(i - 2j + k)^2.$$

($2ij$ and $2jk$ in Fisher's result should be $4ij$ and $4jk$) and on substituting for i, j, k in terms of a, d and m this becomes

$$p^2q^2d^2 = \tfrac{1}{4}\delta^2.$$

The brother–brother correlation is therefore exactly intermediate between parent–offspring correlations with and without the same degree of dominance.

We have set out the above argument in detail in order to show Fisher's procedure. However, the simplest way of finding the above brother–brother table is to use the fact that sibs have probability $\frac{1}{4}$ of sharing two genes in the previously used sense (and therefore of having the same genotype at this locus), probability $\frac{1}{2}$ of sharing one gene, and probability $\frac{1}{4}$ of sharing no gene. The above table is then found by adding the three corresponding 3×3 association tables. The same method of approach can be used in all the following tables but we follow Fisher's method in order to make his discussion clear.

"The effect of dominance is to reduce the fraternal correlation to only half the extent to which the parental correlation is reduced. This allows us to distinguish, as far as the accuracy of the existing figures allows, between the random external effects of environment and those of dominance. This halving of the effect of dominance, it is important to notice, is independent of the relative importance of different factors, of their different degrees of dominance, and of the different proportions in which their phases occur. The correlation between the dominance deviations of siblings is in all cases, $\frac{1}{4}$.

"7. To investigate the cases of uncles and cousins we must deal with all the possible types of mating down to the second generation. The three Mendelian phases will yield six types of mating, and ordinary cousinships are therefore connected by one of six types of sibship. The especially interesting case of double cousins, in which two members of one sibship mate with two members of another, can occur in twenty-one distinct ways, since any pair of the six types of sibship may be taken. The proportionate numbers of the three Mendelian phases in the children produced by the random matings of such pairs of sibships is given in the accompanying table:

Type of sibship ...	1.0.0	1. 1.0	0.1.0	1.2.1	0. 1.1	0.0.1
Frequency ...	p^4	$4p^3q$	$2p^2q^2$	$4p^2q^2$	$4pq^3$	q^4
p^4	1.0.0	3. 1.0	1.1.0	1.1.0	1. 3.0	0.1.0
$4p^3q$	3.1.0	9. 6.1	3.4.1	3.4.1	3.10.3	0.3.1
$2p^2q^2$	1.1.0	3. 4.1	1.2.1	1.2.1	1. 4.3	0.1.1
$4p^2q^2$	1.1.0	3. 4.1	1.2.1	1.2.1	1. 4.3	0.1.1
$4pq^3$	1.3.0	3.10.1	1.4.3	1.4.3	1. 6.9	0.1.3
q^4	0.1.0	0. 3.1	0.1.1	0.1.1	0. 1.3	0.0.1
	$p.q.0$	$\frac{3p}{4} \cdot \frac{p+3q}{4} \cdot \frac{q}{4}$	$\frac{p}{2} \cdot \frac{1}{2} \cdot \frac{q}{2}$	$\frac{p}{2} \cdot \frac{1}{2} \cdot \frac{q}{2}$	$\frac{p}{4} \cdot \frac{3p+q}{4} \cdot \frac{3q}{4}$	$0.p.q$

" The lowest line gives the proportions of the phases in the whole cousinship whose connecting sibship is of each of the six types.

To discuss uncle–nephew relationships and cousins we have to consider three generations because we must first calculate the different probabilities of various classes of sibship which can arise from a random mating of unrelated pairs. This is done in Table F.

TABLE F

Type of mating	Probability of mating	Relative frequency of sibs		
		A_1A_1	A_1A_2	A_2A_2
$A_1A_1 \times A_1A_1$	p^4	1	0	0
$A_1A_1 \times A_1A_2$	$4p^3q$	1	1	0
$A_1A_1 \times A_2A_2$	$2p^2q^2$	0	1	0
$A_1A_2 \times A_1A_2$	$4p^2q^2$	1	2	1
$A_1A_2 \times A_2A_2$	$4pq^3$	0	1	1
$A_2A_2 \times A_2A_2$	q^4	0	0	1

We may illustrate the meaning of this table by saying that the mating $A_1A_2 \times A_1A_2$ has probability $4p^2q^2$ of occurring and that each of its offspring has (independently) the probabilities

$\frac{1}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ of being A_1A_1, A_1A_2 or A_2A_2. Such a sibship is denoted by Fisher by the symbol (1.2.1).

Fisher's 6×6 table is a table giving relative frequencies of the three genetic types in the offspring from a mating in which it is known that one parent comes from one of the above specified sibships and one from another. (Note that the entry 3.10.1 in the fifth row and second column of the 6×6 table should be 3.10.3). Thus the offspring of a mating between an individual out of a sibship whose parental cross was $A_1A_1 \times A_1A_2$, and an individual from a sibship produced by a mating $A_1A_2 \times A_2A_2$, will be of types A_1A_1, A_1A_2 and A_2A_2 with probabilities $\frac{3}{16}$, $\frac{10}{16}$, $\frac{3}{16}$ respectively.

To construct this table it is convenient to regard such symbols as (1, 0, 0), (1, 1, 0), etc., as row vectors. To obtain any entry in the table we premultiply the vector corresponding to the column by the transpose of the vector corresponding to the row. Thus in the above case we take

$$\begin{pmatrix}1\\1\\0\end{pmatrix}(0 \quad 1 \quad 1) = \begin{pmatrix}0 & 1 & 1\\0 & 1 & 1\\0 & 0 & 0\end{pmatrix}.$$

Each element of the resulting 3×3 matrix is then multiplied by the corresponding vector in the following matrix, and the products summed. This matrix is

$$\begin{pmatrix}4.0.0 & 2.2.0 & 0.4.0\\2.2.0 & 1.2.1 & 0.2.2\\0.4.0 & 0.2.2 & 0.0.4\end{pmatrix}.$$

These give relative frequencies of offspring as derived from Table F. Thus the matrix

$$\begin{pmatrix}0 & 1 & 1\\0 & 1 & 1\\0 & 0 & 0\end{pmatrix}$$

gives $$(2.2.0)+(0.4.0)+(1.2.1)+(0.2.2) = (3.10.3)$$

which is the required result.

The table is symmetric about the leading diagonal and has a number of other symmetries.

If an individual from a sibship S_i is mated with an individual chosen at random from the whole population, the three types of individual will occur in the offspring with the probabilities given in the last row. Thus if a member of a sibship of type (1.1.0) is mated in this way, the offspring will be A_1A_1, A_1A_2 and A_2A_2 with probabilities

$$\tfrac{3}{4}p, \tfrac{1}{4}(p+3q), \tfrac{1}{4}q.$$

This can be seen directly or by summing the probabilities corresponding to the elements of the columns of the 6×6 table after multiplying each by the probabilities of the rows, and then rescaling to obtain total probability equal to unity. Thus if two individuals are cousins, and connected by a given one of the above sibships, and are not related in any other way, each will belong to A_1A_1, A_1A_2 or A_2A_2 with probabilities given by the last line.

"If we pick out all possible pairs of uncle (or aunt) and nephew (or niece) we obtain the table

$p^3(p+\frac{1}{2}q)$	$\frac{1}{2}p^2q(3p+q)$	$\frac{1}{2}p^2q^2$
$\frac{1}{2}p^2q(3p+q)$	$\frac{1}{2}pq(p^2+6pq+q^2)$	$\frac{1}{2}pq^2(p+3q)$
$\frac{1}{2}p^2q^2$	$\frac{1}{2}pq^2(p+3q)$	$q^3(\frac{1}{2}p+q)$

the quadratic from which reduces exactly to $\frac{1}{4}\beta^2$, showing that when mating is at random the avuncular correlation is exactly one half of the paternal."

The uncle–nephew table can be constructed from first principles by combining the previous brother–brother table with the parent–offspring table, or it can be constructed from the above 6×6 table. Consider the latter method. Suppose that the uncle is the brother of the nephew's father. There are six sibships in which the father and uncle can occur and these are represented by Fisher by the six row vectors $(1.0.0)$, $(1.1.0)$, $(0.1.0)$, $(1.2.1)$, $(0.1.1)$, $(0.0.1)$ at the top of the six columns of the 6×6 table. The components of these row vectors represent relative frequencies and not probabilities. We therefore convert them into probabilities so that we obtain

$$(1.0.0), \quad (\tfrac{1}{2}.\tfrac{1}{2}.0), \quad (0.1.0), \quad (\tfrac{1}{4}.\tfrac{1}{2}.\tfrac{1}{4}), \quad (0.\tfrac{1}{2}.\tfrac{1}{2}), \quad (0.0.1).$$

These six sibships arise with probabilities

$$p^4, \quad 4p^3q, \quad 2p^2q^2, \quad 4p^2q^2, \quad 4pq^3, \quad q^4$$

respectively and the corresponding probabilities of the A_1A_1, A_1A_2 and A_2A_2 in the nephew are given by the last row of the table. If the vectors of the last row of the table are turned into column vectors $(p.q.0)', \ldots$, etc., the 3×3 association table will have 9 elements which are the elements of the 3×3 matrix

$$\begin{aligned}
&p^4(p.q.0)'(1.0.0) + 4p^3q(\tfrac{3}{4}p.\tfrac{1}{4}(p+3q).\tfrac{1}{4}q)'(\tfrac{1}{2}.\tfrac{1}{2}.0) + 2p^2q^2(\tfrac{1}{2}p.\tfrac{1}{2}.\tfrac{1}{2}q)'(0.1.0) \\
&\quad + 4p^2q^2(\tfrac{1}{2}p.\tfrac{1}{2}.\tfrac{1}{2}q)'(\tfrac{1}{4}.\tfrac{1}{2}.\tfrac{1}{4}) + 4pq^3(\tfrac{1}{4}p.\tfrac{1}{4}(3p+q).\tfrac{3}{4}q)'(0.\tfrac{1}{2}.\tfrac{1}{2}) + q^4(0.p.q)'(0.0.1) \\
&= p^4\begin{pmatrix} p & 0 & 0 \\ q & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 2p^3q\begin{pmatrix} \frac{3}{4}p & \frac{3}{4}p & 0 \\ \frac{1}{4}(p+3q) & \frac{1}{4}(p+3q) & 0 \\ \frac{1}{4}q & \frac{1}{4}q & 0 \end{pmatrix} \\
&\quad + 2p^2q^2\begin{pmatrix} 0 & \frac{1}{2}p & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2}q & 0 \end{pmatrix} + p^2q^2\begin{pmatrix} \frac{1}{2}p & p & \frac{1}{2}p \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{2}q & q & \frac{1}{2}q \end{pmatrix} \\
&\quad + 2pq^3\begin{pmatrix} 0 & \frac{1}{4}p & \frac{1}{4}p \\ 0 & \frac{1}{4}(3p+q) & \frac{1}{4}(3p+q) \\ 0 & \frac{3}{4}q & \frac{3}{4}q \end{pmatrix} + q^4\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & p \\ 0 & 0 & q \end{pmatrix}
\end{aligned}$$

and adding these we obtain the uncle–nephew table given by Fisher. Notice that this table is symmetric although the relationship is not. The rows correspond to the nephew and the columns to the uncle. Inserting the values i, j, k and multiplying each element of the matrix by the corresponding product of i's, j's and k's we get a formula for the covariance which begins

$$p^3(p+\tfrac{1}{2}q)\,i^2 + 2\{\tfrac{1}{2}p^2q(3p+q)\,ij\} + \ldots.$$

Substituting for $j = -\dfrac{ip^2+kq^2}{2pq}$ the covariance reduces to

$$\frac{1}{8pq}(p^2i - q^2k)^2 = \tfrac{1}{4}\beta^2.$$

Thus there is no correlation due to dominance.

"From the twenty-one types of double cousinship pairs may be picked, the proportions of which are shown in the table:

$p^2(p+\frac{1}{4}q)^2$	$\frac{3}{2}p^2q(p+\frac{1}{4}q)$	$\frac{9}{16}p^2q^2$
$\frac{3}{2}p^2q(p+\frac{1}{4}q)$	$\frac{1}{2}pq(p^2+\frac{3}{2}pq+q^2)$	$\frac{3}{2}pq^2(\frac{1}{4}p+q)$
$\frac{9}{16}p^2q^2$	$\frac{3}{2}pq^2(\frac{1}{4}p+q)$	$q^2(\frac{1}{4}p+q)^2$

which agrees with the table given by Snow for ordinary first cousins. I cannot explain this divergence, unless it be that Snow is in error, my values for ordinary first cousins leading to less than half this value for the

correlation. Simplifying the quadratic in i, j, k, which is most easily done in this case by comparison with the avuncular table, we find for the correlation of double cousins

$$\frac{1}{4\sigma^2}(\tau^2+\tfrac{1}{4}\epsilon^2),$$

showing that double cousins, like brothers, show some similarity in the distribution of deviations due to dominance, and that with these cousins the correlation will in general be rather higher than it is for uncle and nephew."

Double cousinship is more complicated. Suppose the cousins are such that the two fathers come from one sibship and the two mothers from another. There are therefore 36 possibilities of which it is only necessary to consider 21 by symmetry. In the 6×6 table the individual entries are 3 element vectors whose components are proportional to the frequencies of A_1A_1, A_1A_2 and A_2A_2 in the progeny of a mating between individuals chosen from these sibships. Double cousins are the results of independent choice of pairs from the same sibships in this way. We can therefore construct a 6×6 table (Table G) in each cell of which we have first the probability that the two

TABLE G

p^8	$4p^7q$	$2p^6q^2$	$4p^6q^2$	$4p^5q^3$	p^4q^4
1 0 0 0 0 0 0 0 0	9 3 0 3 1 0 0 0 0	1 1 0 1 1 0 0 0 0	1 1 0 1 1 0 0 0 0	1 3 0 3 9 0 0 0 0	0 0 0 0 1 0 0 0 0

$16p^6q^2$	$8p^5q^3$	$16p^5q^3$	$16p^4q^4$	$4p^3q^5$
81 54 9 54 36 6 9 6 1	9 12 3 12 16 4 3 4 1	9 12 3 12 16 4 3 4 1	9 30 9 30 100 30 9 30 9	0 0 0 0 9 3 0 3 1

$4p^4q^4$	$8p^4q^4$	$8p^3q^5$	$2p^2q^6$
1 2 1 2 4 2 1 2 1	1 2 1 2 4 2 1 2 1	1 4 3 4 16 12 3 12 9	0 0 0 0 1 1 0 1 1

$16p^4q^4$	$16p^3q^5$	$4p^2q^6$
1 2 1 2 4 2 1 2 1	1 4 3 4 16 12 3 12 9	0 0 0 0 1 1 0 1 1

$16p^2q^6$	$4pq^7$
1 6 9 6 36 54 9 54 81	0 0 0 0 1 3 0 3 9

q^8
0 0 0 0 0 0 0 0 1

corresponding sibships have been chosen and then a 3×3 matrix whose elements are proportional to the probabilities of the three genetic phases in the two double cousins. It is more convenient to enter the matrix with numbers which are only proportional to the probabilities and not equal to them, as in this way we avoid the use of fractions. To obtain the probabilities it is necessary to divide each element by the sum of all the elements in the matrix.

Thus if one of the connecting sibships corresponds to the symbol (0.1.0) and the other to (3.1.0), the vector given in Fisher's 6×6 table is (3.4.1), and the contribution to the covariance table will have elements proportional to

$$(3.4.1)'(3.4.1) = \begin{pmatrix} 9 & 12 & 3 \\ 12 & 16 & 4 \\ 3 & 4 & 1 \end{pmatrix},$$

and since the corresponding probability is

$$(4p^3q) \times (2p^2q^2) = 8p^5q^3,$$

the sum of the elements of the matrix is 64, and there is another equal contribution from the matrix situated symmetrically on the other side of the main diagonal, the contribution to the covariance table is

$$\tfrac{1}{4}p^5q^2 \begin{pmatrix} 9 & 12 & 3 \\ 12 & 16 & 4 \\ 3 & 4 & 1 \end{pmatrix}.$$

The empty cells are obtained by symmetry. Multiplying by the probabilities, the reciprocal of the sum of the elements of each matrix, and adding, we check Fisher's table for double cousins. The difference of this table from the uncle–nephew table is

$$\begin{matrix} \tfrac{1}{16}p^2q^2 & -\tfrac{1}{8}p^2q^2 & \tfrac{1}{16}p^2q^2 \\ -\tfrac{1}{8}p^2q^2 & \tfrac{1}{4}p^2q^2 & -\tfrac{1}{8}p^2q^2 \\ \tfrac{1}{16}p^2q^2 & -\tfrac{1}{8}p^2q^2 & \tfrac{1}{16}p^2q^2 \end{matrix}$$

which gives a term $\tfrac{1}{16}p^2q^2(i-2j+k)^2 = \tfrac{1}{16}\delta^2$,

and the correlation of double cousins is therefore

$$\frac{1}{4\sigma^2}(\tau^2 + \tfrac{1}{4}\epsilon^2).$$

Notice that the double cousin table is necessarily symmetric.

" For ordinary first cousins I find the following table of the distribution of random pairs drawn from the six types of ordinary cousinship:

$$\begin{matrix} \tfrac{1}{4}p^3(4p+q) & \tfrac{1}{4}p^2q^2(7p+q) & \tfrac{3}{4}p^2q^2 \\ \tfrac{1}{4}p^2q(7p+q) & \tfrac{1}{4}pq(p^2+14pq+q^2) & \tfrac{1}{4}pq^2(p+7q) \\ \tfrac{3}{4}p^2q^2 & \tfrac{1}{4}pq^2(p+7q) & \tfrac{1}{4}q^3(p+4q) \end{matrix}$$

which yields the correlation $\frac{1}{8}\frac{\tau^2}{\sigma^2}$. "

Ordinary first cousins are connected by a single sibship. They are therefore each the result of the mating of one of the sibships in Fisher's 6×6 table with a mate chosen at random. We can therefore divide all first cousins into 6 classes according to the type of connecting sibship and the covariance table is the sum of 6 tables. Each of the latter is obtained by multiplying the probabilities of the connecting sibship by the matrix obtained by the column into row product of the last row of Fisher's 6×6 table by itself. Thus for example the first of these is $(p.q.0)$ with probability p^4, and its contribution is

$$p^4(p.q.0)'(p.q.0) = p^4 \begin{pmatrix} p^2 & pq & 0 \\ pq & q^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The sum of all these is

$$p^4\begin{pmatrix} p^2 & pq & 0 \\ pq & q^2 & 0 \\ 0 & 0 & 0 \end{pmatrix} + 4p^3q\begin{pmatrix} \frac{9}{16}p^2 & \frac{1}{16}(3p^2+9pq) & \frac{3}{16}pq \\ \frac{1}{16}(3p^2+9pq) & \frac{1}{16}(p+3q)^2 & \frac{1}{16}(pq+3q^2) \\ \frac{3}{16}pq & \frac{1}{16}(pq+3q^2) & \frac{1}{16}q^2 \end{pmatrix}$$

$$+ 6p^2q^2\begin{pmatrix} \frac{1}{4}p^2 & \frac{1}{4}p & \frac{1}{4}pq \\ \frac{1}{4}p & \frac{1}{4} & \frac{1}{4}q \\ \frac{1}{4}pq & \frac{1}{4}q & \frac{1}{4}q^2 \end{pmatrix} + 4pq^3\begin{pmatrix} \frac{1}{16}p^2 & \frac{1}{16}(3p^2+pq) & \frac{3}{16}pq \\ \frac{1}{16}(3p^2+pq) & \frac{1}{16}(3p+q)^2 & \frac{1}{16}(9pq+3q^2) \\ \frac{3}{16}pq & \frac{1}{16}(9pq+3q^2) & \frac{9}{16}q^2 \end{pmatrix} + q^4\begin{pmatrix} 0 & 0 & 0 \\ 0 & p^2 & pq \\ 0 & pq & q^2 \end{pmatrix}$$

Adding these we obtain Fisher's cousin table which is checked except for the entry in the first row and second column which should be

$$\tfrac{1}{4}p^2q(7p+q).$$

This table is necessarily symmetric.

Calculating the covariance and subtracting

$$(p^2i+2pqj+q^2k)^2 = 0,$$

we get

$$\tfrac{1}{4}pq(pi-(p-q)j+qk)^2 = \frac{1}{16pq}(p^2i-q^2k)^2 = \tfrac{1}{8}\beta^2,$$

so that for single cousins there is no dominance component in the correlation.

" In a similar way the more distant kin may be investigated, but since for them reliable data have not yet been published, the table already given of genetic correlations will be a sufficient guide.

" 8. Before extending the above results to the more difficult conditions of assortative mating, it is desirable to show how our methods may be developed so as to include the statistical feature to which we have applied the term Epistacy. The combination of two Mendelian factors gives rise to nine distinct phases, and there is no biological reason for supposing that nine such distinct measurements should be exactly represented by the nine deviations formed by adding i, j, or k to i', j', or k'. If we suppose that i, j, k, i', j', k' have been so chosen as to represent the nine actual types with the least square error, we have now to deal with additional quantities, which we may term

$$\begin{matrix} e_{11} & e_{12} & e_{13}, \\ e_{21} & e_{22} & e_{23}, \\ e_{31} & e_{32} & e_{33}, \end{matrix}$$

connected by the six equations, five of which are independent,

$$\begin{aligned} p^2e_{11}+2pqe_{21}+q^2e_{31} &= 0, & p'^2e_{11}+2p'q'e_{12}+q'^2e_{13} &= 0, \\ p^2e_{12}+2pqe_{22}+q^2e_{32} &= 0, & p'^2e_{21}+2p'q'e_{22}+q'^2e_{23} &= 0, \\ p^2e_{13}+2pqe_{23}+q^2e_{33} &= 0, & p'^2e_{31}+2p'q'e_{32}+q'^2e_{33} &= 0." \end{aligned}$$

The definitions of i, j, k are now modified. Suppose that we have two non-linked loci at which the genotypes are A_1A_1, A_1A_2, A_2A_2 and B_1B_1, B_1B_2, B_2B_2. Let their effect in combination be $a_{11}, \ldots$ so we have Table H.

TABLE H

		B_1B_1 i'	B_1B_2 j'	B_2B_2 k'
A_1A_1	i	a_{11}	a_{12}	a_{13}
A_1A_2	j	a_{21}	a_{22}	a_{23}
A_2A_2	k	a_{31}	a_{32}	a_{33}

We assume random mating and put p, p', for the frequencies of A_1 and B_1 respectively. We also assume that random mating has been occurring in the population for a sufficiently long time for

the frequencies of the genotypes $A_1A_1B_1B_1$, $A_1A_2B_1B_1$, etc., to have attained their limiting values $p^2p'^2$, $(2pq)p'^2$, etc.

The values i, j, k, i', j', k', are now chosen so the sum of the corresponding values for the two loci represent a_{11}, a_{12}, ... as closely as possible in the sense of least squares. Write

$$\begin{aligned} e_{11} &= a_{11}-i-i', & e_{23} &= a_{23}-j-k', \\ e_{12} &= a_{12}-i-j', & e_{31} &= a_{31}-k-i', \\ e_{13} &= a_{13}-i-k', & e_{32} &= a_{32}-k-j', \\ e_{21} &= a_{21}-j-i', & e_{33} &= a_{33}-k-k'. \\ e_{22} &= a_{22}-j-j', \end{aligned}$$

Then we want to minimize the sum

$$S_1 = p^2q'^2e_{11}^2+2p^2p'q'e_{12}^2+p^2q'^2e_{13}^2+2pqp'^2e_{21}^2+4pqp'q'e_{22}^2+2pqq'^2e_{23}^2 + q^2p'^2e_{31}^2+2q^2p'q'e_{32}^2+q^2q'^2e_{33}^2.$$

Differentiating S_1 with respect to i, j, k and i', j', k' we get the six equations given above by Fisher. This process is exactly analogous to estimating row and column effects in an experiment in which rows and columns are orthogonal, the orthogonality being here a consequence of independent distribution of the two factors. Since the addition of a constant to all the a's makes no difference, we can choose the latter so that the mean of i, j, k and the mean of the i', j', k' are both zero. This means that

$$p^2i+2pqj+q^2k = 0, \quad p'^2i'+2p'q'j'+q'^2k' = 0. \tag{XII a}$$

Of the six equations only five can be independent in general for if we multiply the first three by p'^2, $2p'q'$, q'^2, respectively and add we get the same result as multiplying the second three by p^2, $2pq$, q^2, and adding. Thus only $4 = 9-5$ of the e's can be varied and the epistatic and dominance relations arising from two different factors require four constants for their definition.

"This is a complete representation of any such deviations from linearity as may exist between two factors. Such dual epistacy, as we may term it, is the only kind of which we shall treat. More complex connections could doubtless exist, but the number of unknowns introduced by dual epistacy alone, four, is more than can be determined by existing data. In addition it is very improbable that any statistical effect, of a nature other than that which we are considering, is actually produced by more complex somatic connections.

The full association table between two relatives, when we are considering two distinct Mendelian factors, consists of eighty-one cells, and the quadratic expression to which it leads now involves the nine epistatic deviations. A remarkable simplification is, however, possible, since each quantity, such as e_{21}, which refers to a partially or wholly heterozygous individual, is related to two other quantities, such as e_{11} and e_{31}, by just the same equation as that by which j is related to i and k, and occurs in the 9×9 table with corresponding coefficients. The elimination of the five deviations e_{21}, e_{12}, e_{32}, e_{23}, e_{22} is therefore effected by rewriting the 9×9 table as a 4×4 table, derived from the quadratic in i and k corresponding to the relationship considered."

By the method of definition we can split the contribution to the total character due to the genes at the A and B loci into seven components,

$$x = x_1+x_2+x_3+x_1'+x_2'+x_3'+x_4'',$$

where x_1 is the effect due to the 'first' gene at the A locus (e.g. that inherited from father), x_2 the effect due to the second gene at the first locus (e.g. that from mother), and x_3 is the deviation from linearity due to dominance. Thus $x_1+x_2+x_3 = i, j, k$ according as the genotype is A_1A_1, A_1A_2, or

A_2A_2. x_1', x_2', x_3' are the corresponding components at the second locus, and x_4'' is the 'epistacy' deviation.

We have already seen that x_1, x_2, x_3, are uncorrelated in pairs, and the same holds good for x_1', x_2', x_3'. The latter are also statistically independent of x_1, x_2, x_3 because they are at an unlinked locus.

It also follows from the set of equations such as

$$p^2e_{11}+2pqe_{21}+q^2e_{31}=0,$$

that for any fixed genotype, such as B_1B_1, at the second locus, the mean value of $x_4''(=e_{rs})$ is zero. Thus x_4'' is uncorrelated with the effects x_1', x_2', x_3', at the second locus. A similar argument holds for the first locus. This could also have been seen from Least Square Regression Theory.

The seven components above are therefore uncorrelated between themselves, and the variance of x decomposes into seven orthogonal components,

$$\operatorname{var}x=\Sigma\operatorname{var}(x_i)+\Sigma\operatorname{var}(x_i')+\operatorname{var}(x_4'').$$

We can similarly write the value, X, of the character for a relative as

$$X=X_1+X_2+X_3+X_1'+X_2'+X_3'+X_4'',$$

where the X_i have been numbered in such a way that any genes shared by the two relatives will affect only x_r and X_r (or x_r' and X_r') with the same suffix. Then because of this numbering and the previous results,

$$\operatorname{cov}(x,X)=\Sigma\operatorname{cov}(x_r,X_r)+\Sigma\operatorname{cov}(x_r',X_r')+\operatorname{cov}(x_4'',X_4'').$$

The only terms that therefore remain to be considered are $\operatorname{var}(x_4'')$, $\operatorname{var}(X_4'')$, and $\operatorname{cov}(x_4'',X_4'')$.

$\operatorname{var}(x_4'')$ can be written

$$p'^2\{p^2e_{11}^2+2pqe_{21}^2+q^2e_{31}^2\}+2p'q'\{p^2e_{12}^2+2pqe_{22}^2+q^2e_{32}^2\}+q'^2\{p^2e_{13}^2+2pqe_{23}^2+q^2e_{33}^2\}$$
$$=p'^2A+2p'q'B+q'^2C,\quad\text{say.}$$

We can treat each of the quadratic forms A, B, C in the same way. Consider the first. This is a quadratic form in e_{11}, e_{21}, e_{31} whose matrix is

$$\begin{pmatrix}p^2&0&0\\0&2pq&0\\0&0&q^2\end{pmatrix},$$

where the rows correspond to e_{11}, e_{21}, e_{31} and the columns also to e_{11}, e_{21}, e_{31}. We turn this into a quadratic form in e_{11}, e_{31} only by using the relation

$$e_{21}=\frac{1}{2pq}(-p^2e_{11}-q^2e_{31}),$$

which is one of the equations derived by least squares. The quadratic form A then becomes

$$p^2e_{11}^2+2pq\left\{\frac{p^2e_{11}+q^2e_{31}}{2pq}\right\}^2+q^2e_{31}^2=\left(p^2+\frac{p^3}{2q}\right)e_{11}^2+pqe_{11}e_{31}+\left(q^2+\frac{q^3}{2p}\right)e_{31}^2,$$

which has the 2×2 matrix

$$T_1=\begin{pmatrix}\frac{p^2}{2q}(p+2q)&\frac{1}{2}pq\\\frac{1}{2}pq&\frac{q^2}{2p}(2p+q)\end{pmatrix},$$

where the rows correspond to e_{11}, e_{31} respectively, and the columns similarly. Exactly the same relationships apply to B and C resulting in the same 2×2 matrix. We can therefore write

$$\begin{aligned}\text{var}(x_4'') &= \left(p^2+\frac{p^3}{2q}\right)\{p'^2e_{11}^2+2p'q'e_{12}^2+q'^2e_{13}^2\}+pq\{p'^2e_{11}e_{31}+2p'q'e_{12}e_{32}+q'^2e_{13}e_{33}\}\\ &\quad+\left(q^2+\frac{q^3}{2p}\right)\{p'^2e_{31}^2+2p'q'e_{32}^2+q'^2e_{33}^2\}\\ &=\left(p^2+\frac{p^3}{2q}\right)D+pqE+\left(q^2+\frac{q^3}{2p}\right)F, \quad \text{say.}\end{aligned}$$

We now apply the same procedure to D, E, F. D becomes

$$\left(p'^2+\frac{p'^3}{2q'}\right)e_{11}^2+p'q'e_{11}e_{13}+\left(q'^2+\frac{q'^3}{2p'}\right)e_{13}^2,$$

E becomes $$\left(p'^2+\frac{p'^3}{2q'}\right)e_{11}e_{31}+\tfrac{1}{2}p'q'e_{11}e_{33}+\tfrac{1}{2}p'q'e_{13}e_{31}+\left(q'^2+\frac{q'^3}{2p'}\right)e_{13}e_{33},$$

and F becomes $$\left(p'^2+\frac{p'^3}{2q'}\right)e_{31}^2+p'q'e_{31}e_{33}+\left(q'^2+\frac{p'^3}{2q'}\right)e_{33}^2.$$

We have therefore reduced the quadratic form in the nine variables $e_{11}, \ldots, e_{33}$ to one in four variables e_{11}, e_{13}, e_{31}, e_{33}. This is

$$\begin{aligned}&\left(p^2+\frac{p^3}{2q}\right)\left\{\left(p'^2+\frac{p'^3}{2q'}\right)e_{11}^2+p'q'e_{11}e_{13}+\left(q'^2+\frac{q'^3}{2p'}\right)e_{13}^2\right\}\\ &\quad+pq\left\{\left(p'^2+\frac{p'^3}{2q'}\right)e_{11}e_{31}+\tfrac{1}{2}p'q'e_{11}e_{33}+\tfrac{1}{2}p'q'e_{13}e_{31}+\left(q'^2+\frac{q'^3}{2p'}\right)e_{13}e_{33}\right\}\\ &\quad+\left(q^2+\frac{q^3}{2p}\right)\left\{\left(p'^2+\frac{p'^3}{2q'}\right)e_{31}^2+p'q'e_{31}e_{33}+\left(q'^2+\frac{q'^3}{2p'}\right)e_{33}^2\right\}.\end{aligned}$$

On taking out a factor $1/4pqp'q'$ this agrees with Fisher's result.

The method of deriving the above can be described in algebraic form. Given the matrix T_1 above and the corresponding matrix T_2 corresponding to p, q, we form the direct product, which is of order 4, and is obtained by replacing each element of T_2 by the product of this element considered as a scalar with the matrix T_1.

The second expression below for this quadratic is verified by expanding all the above terms and subtracting the terms obtained from the expansion of

$$(p^2p'^2e_{11}-p^2q'^2e_{13}-q^2p'^2e_{31}+q^2q'^2e_{33})^2,$$

whence we obtain four similar expressions of which a typical one is

$$2pqp'^3q'(pe_{11}+qe_{31})^2+2pqp'^4(pe_{11}+qe_{31})^2=2pqp'^3(pe_{11}+qe_{31})^2,$$

because $p'+q'=1$. Thus the four similar terms in Fisher's second expression below are correct although they are only of the seventh degree in p, q, p', q', and not of the eighth as in the original quadratic form.

" Thus the variance, found by squaring the individual variations, is derived from the 3×3 table

p^2	—	—
—	$2pq$	—
—	—	q^2

which yields the 2×2 table

$$\begin{array}{cc} \frac{p^2}{2q}(p+2q) & \frac{1}{2}pq \\ \frac{1}{2}pq & \frac{q^2}{2p}(2p+q) \end{array}$$

and the quadratic in e_{11}, e_{13}, e_{31}, e_{33}

$$\frac{1}{4pqp'q'}[(p+2q)(p'+2q')p^3p'^3 e_{11}^2 + 3 \text{ similar terms} + 2p^2q^2p'^3(p'+2q')e_{11}e_{31} + 3 \text{ similar terms} + 2p^2q^2p'^2q'^2(e_{11}e_{33}+e_{13}e_{31})],$$

which also takes the form

$$\frac{1}{4pqp'q'}[(p^2p'^2e_{11} - p^2q'^2e_{13} - q^2p'^2e_{31} + q^2q'^2e_{33})^2 + 2pqp'^3(pe_{11}+qe_{31})^2 + 3 \text{ similar terms}].$$

The parental table

$$\begin{array}{cc} p^3/4q & -\frac{1}{4}pq \\ -\frac{1}{4}pq & q^3/4p \end{array}$$

yields

$$\frac{1}{16pqp'q'}[p^2p'^2e_{11} - p^2q'^2e_{13} - q^2p'^2e_{31} + q^2q'^2e_{33}]^2.$$

The parent–offspring table has $9 \times 9 = 81$ cells. In order to reduce it we use the same kind of transformations as in discussing the variance. Consider terms of the form

$$e_{rm}e_{sn},$$

where m, n are fixed and r, $s = 1, 2, 3$. From the parent–offspring table (Table C above) with $P = p^2$, $Q = pq$, $R = q^2$ we obtain the following similar table (Table I), using the fact that the two loci segregate independently.

TABLE I

	Parent		
Offspring	e_{1n}	e_{2n}	e_{3n}
e_{1m}	p^3	p^2q	0
e_{2m}	p^2q	pq	pq^2
e_{3m}	0	pq^2	q^3

We now use the formulae:

$$e_{2m} = -\frac{1}{2pq}(p^2e_{1m} + q^2e_{3m}), \quad e_{2n} = -\frac{1}{2pq}(p^2e_{1n} + q^2e_{3n}).$$

We turn the quadratic form above (which has 6 variables if $m \neq n$ and 3 if $m = n$) into a quadratic form with 4 variables if $m \neq n$ and 2 variables if $m = n$. We then get the array given in Table J,

TABLE J

	e_{1n}	e_{2n}	e_{3n}
e_{1m}	$p^3/4q$	0	$-\frac{1}{4}pq$
e_{2m}	0	0	0
e_{3m}	$-\frac{1}{4}pq$	0	$q^3/4p$

which we can regard as a 2×2 table. Using this and the similar table with p', q', the direct product of the two matrices gives an array whose elements are the elements of Table K multiplied by $\frac{1}{16pqp'q'}$. The corresponding quadratic form is obviously

$$\frac{1}{16pqp'q'}\{p^2p'^2e_{11} - p^2q'^2e_{13} - q^2p'^2e_{31} + q^2q'^2e_{33}\}^2.$$

TABLE K

	e_{11}	e_{31}	e_{13}	e_{33}
e_{11}	$p^4p'^4$	$-p^2q^2p'^4$	$-p^4p'^2q'^2$	$p^2q^2p'^2q'^2$
e_{31}	$-p^2q^2p'^4$	$q^4p'^4$	$p^2q^2p'^2q'^2$	$-q^4p'^2q'^2$
e_{13}	$-p^4p'^2q'^2$	$p^2q^2p'^2q'^2$	$p^4q'^4$	$-p^2q^2q'^4$
e_{33}	$p^2q^2p'^2q'^2$	$-q^4p'^2q'^2$	$-p^2q^2q'^4$	$q^4q'^4$

" and the fraternal table

$p^2/4q$	—
—	$q^2/4p$

leads us to the simple expression

$$\frac{1}{16pqp'q'}[p^3p'^3e_{11}^2+p^3q'^3e_{13}^2+q^3p'^3e_{31}^2+q^3q'^3e_{33}^2]."$$

Applying the same argument to the brother–brother table (Table E) and eliminating e_{2m} and e_{2n} from the corresponding quadratic form we get the array of Table L, which combined with the similar result for p', q' gives

$$\frac{1}{16pqp'q'}\{p^3p'^3e_{11}^2+p^3q'^3e_{13}^2+q^3p'^3e_{31}^2+q^3q'^3e_{33}\}.$$

TABLE L

	e_{1n}	e_{3n}
e_{1m}	$p^2/4q$	0
e_{3m}	0	$q^2/4p$

" For uncles and cousins we obtain respectively $\frac{1}{4}$ and $\frac{1}{16}$ of the parental contribution, while for double cousins the table

$\frac{p^2}{16q}(2p+q)$	$-\frac{1}{16}pq$
$-\frac{1}{16}pq$	$\frac{q^2}{16p}(p+2q)$

and a quadratic similar to that for the variance."

The same technique is applied to the uncle–nephew table to give Table M.

TABLE M

	e_{1n}	e_{3n}
e_{1m}	$p^3/8q$	$-\frac{1}{8}pq$
e_{3m}	$-\frac{1}{8}pq$	$q^3/8p$

Since this is one-half of the corresponding table for parent–offspring, the epistatic contribution to the covariance between uncle and nephew is $\frac{1}{4}$ that of the epistatic component in parent–offspring covariance.

Cousins and double cousins are then easily treated in the same way.

" 9. With assortative mating all these coefficients will be modified. There will be association between similar phases of different factors, so that they cannot be treated separately. There will also be an increase in the variance.

" We must determine the nature of the association between different factors, and ascertain how it is related to the degree of assortative mating necessary to maintain it. Then we shall be able to investigate the statistical effects of this association on the variance of the population and on the correlations.

" If μ be the marital correlation, then in a population with variance V the frequency of individuals in the range dx is

$$\frac{1}{\sqrt{(2\pi V)}} e^{-x^2/2V}\,dx = M,$$

and the frequency in the range dy is $\quad \frac{1}{\sqrt{(2\pi V)}} e^{-y^2/2V}\,dy = N;$

but the frequency of matings between these two groups is not simply MN, as would be the case if there were no marital correlation, but

$$\frac{1}{2\pi V\sqrt{(1-\mu^2)}} \exp\left\{-\frac{1}{1-\mu^2}\,\frac{x^2-2\mu xy+y^2}{2V}\right\} dx\,dy,$$

which is equal to

$$\frac{MN}{\sqrt{(1-\mu^2)}} \exp\left\{-\frac{\mu^2x^2-2\mu xy+\mu^2y^2}{2V(1-\mu^2)}\right\}.$$

" In studying the effect of assortative mating we shall require to know the frequency of matings between two groups, each with a variance nearly equal to that of the whole population, but centred about means a and b. The frequencies of such groups in any ranges dx, dy can be written down, and if the chance of any mating depends only on x and y, the frequency of mating between these two groups can be expressed as a double integral. If M and N are the frequencies in the two groups, the frequency of mating between them is found to be

$$MN\,e^{\mu ab/V}."$$

The idea in the above section is that non-randomness in mating is due to a tendency for the biometric measurements in the two mating individuals to be correlated. Suppose that these two measurements are x and y, and that since they are the result of a large number of independently segregating factors they can be supposed to be normally distributed. It is assumed that there is no epistasis. We take their means as zero and their variances as V. The probabilities that they lie respectively in ranges $(x, x+dx)$ and $(y, y+dy)$ are taken as M and N, and it is assumed that their joint probability distribution is given by the bivariate normal distribution above with μ as correlation coefficient.

The expression

$$(1-\mu^2)^{-\frac{1}{2}} \exp\left\{-\frac{\mu^2x^2-2\mu xy+\mu^2y^2}{2V(1-\mu^2)}\right\}$$

can be regarded as a weighting factor giving the relative probability of a mating between two particular individuals which are known to have the measurements x and y.

Now suppose that these two individuals are chosen at random out of normal populations which are known to have the means a and b respectively and variances equal to V. Their relative probability of mating is given by

$$\frac{1}{2\pi V(1-\mu^2)^{\frac{1}{2}}} \iint \exp\left\{-\frac{(x-a)^2}{2V}-\frac{(y-b)^2}{2V}-\frac{\mu^2x^2-2\mu xy+\mu^2y^2}{2V(1-\mu)^2}\right\} dx\,dy$$

$$= \frac{1}{2\pi V(1-\mu^2)^{\frac{1}{2}}} \iint \exp\left\{-\frac{W}{2V(1-\mu^2)}\right\} dx\,dy,$$

where $\quad W = (x-m_1)^2 - 2\mu(x-m_1)(y-m_2) + (y-m_2)^2 + K,$

where $\quad m_1 = a+\mu b,\ m_2 = b+\mu a \quad \text{and} \quad K = -2\mu ab(1-\mu^2).$

Integrating out, the expression becomes

$$\exp\frac{\mu ab}{V}$$

as required. Notice that this result is exact so long as the bivariate distribution is truly normal.

When $\mu \neq 0$ the non-randomness of the mating has two effects: (1) The Hardy–Weinberg equilibrium for each individual locus is destroyed. (2) The zygotic frequencies for different factors are no longer independent. Hence the average values of individuals of the form A_1A_1, A_1A_2 and A_2A_2 cannot be taken as i, j and k, but depend also on other loci. It is this which introduces the essential complication and requires the introduction of the condition that the population is stationary.

Fisher now investigates the effect of assortative mating on the genotype frequencies, using the condition that these frequencies are the same in the offspring generation as in the parent generation. This implies that the probability of being a parent is independent of genotype so that there are no selective differences. It is possible to devise schemes of assortative mating in which, for example, the extreme types are less likely to find suitable mates. In such a case the distribution amongst the offspring of all the matings would be that of the population as a whole but not the same as that amongst 'parents'.

We first consider the effect on the frequencies of the three phases of a single factor. Write D, H and R for A_1A_1, A_1A_2 and A_2A_2. Consider the effects of the various types of mating listed below:

Mating	Offspring	Mating	Offspring
$D \times D$	D	$H \times H$	$\frac{1}{4}D + \frac{1}{2}H + \frac{1}{4}R$
$D \times H$	$\frac{1}{2}D + \frac{1}{2}H$	$H \times R$	$\frac{1}{2}H + \frac{1}{2}R$
$D \times R$	H	$R \times R$	R

The first two and last two of these matings will, in an indefinitely large population, produce no change in zygotic frequency since the relative proportions of the phases in the offspring are the same as those of the parents.

Out of all possible matings let the frequencies of mating $D \times R$ and $H \times H$ be f_1 and f_2 respectively. Then the contribution of these matings to the next generation will be such that D, H, R are in the proportion

$$\tfrac{1}{4}f_2, \quad f_1 + \tfrac{1}{2}f_2, \quad \tfrac{1}{4}f_2,$$

whilst the proportion of D, H and R amongst the mates entering into these matings is $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$. If these ratios are to be the same we must have $f_2 = 2f_1$. Let I, J, K be the means of the character in the individuals which are D, H and R. Since there are supposed to be many loci contributing to the character, the contribution of any one locus to the whole character is small, so that

$$I/V^{\frac{1}{2}}, \quad J/V^{\frac{1}{2}}, \quad K/V^{\frac{1}{2}}$$

are all small and the variance of the character for a given phase at this locus is practically V. Hence the frequencies of these matings are proportional to

$$e^{\mu IK/V} \quad \text{and} \quad e^{\mu J^2/V},$$

by the above theory, and IK/V, J^2/K are quantities of the second order of smallness. Hence to a high degree of approximation $4Q^2 e^{\mu J^2/V} = 4PR\, e^{\mu IK/V}$.

Expanding the exponentials and neglecting squares of J^2/V and IK/V, we get

$$PR - Q^2 = (\mu/V)\{Q^2J^2 - PRIK\} = \mu(Q^2/V)\{J^2 - IK\},$$

on observing that $PR - Q^2$ is of the second order of smallness. Note that this is an approximation. We now put

$$P = p^2 + \delta_1, \quad Q = pq + \delta_2, \quad R = q^2 + \delta_3.$$

Then $\delta_1 + 2\delta_2 + \delta_3 = 0$. The gene frequency of A_1 must be

$$p = P + Q = p^2 + pq + \delta_1 + \delta_2 = p + \delta_1 + \delta_2$$

and hence $\delta_1 + \delta_2 = 0$. Similarly, $\delta_2 + \delta_3 = 0$. Hence

$$\delta_1 = -\delta_2 = \delta_3 = \delta, \quad \text{say.}$$

If (XIII) is to be satisfied with $(J^2 - IK)\, V^{-1}$ small, δ must be small also, and substituting we get

$$(p^2 + \delta)(q^2 + \delta) - (pq - \delta)^2 = (pq - \delta)^2 \mu (J^2 - IK)/V,$$

and to a first approximation $\delta = p^2 q^2 \mu (J^2 - IK)/V$,

so that (XIV) follows.

The deviation in P, Q, R, from the values they would have if the Hardy–Weinberg equation held, are of the second order of smallness when

$$I/V^{\frac{1}{2}}, \quad J/V^{\frac{1}{2}}, \quad K/V^{\frac{1}{2}}$$

are regarded as being of the first order of smallness. Equation (XV) follows from the definition of I, J, K and we can use this to eliminate J. The first approximation to J is got by putting p^2, pq, q^2 for P, Q, R so that

$$J = \frac{1}{2pq}(p^2 I + q^2 K),$$

and putting this in the expression for α we get

$$\frac{\mu}{V}\{\tfrac{1}{4}(p^2 I + q^2 K)^2 - p^2 q^2 IK\} = \frac{\mu}{4V}(p^2 I - q^2 K)^2.$$

" 10. We shall apply this expression first to determine the equilibrium value of the frequencies of the three phases of a single factor. Of the six types of mating which are possible, all save two yield offspring of the same genetic phase as their parents. With the inbreeding of the pure forms $D \times D$ and $R \times R$ obviously no change is made, and the same is true of the crosses $D \times H$ and $R \times H$, for each of these yields the pure form and the heterozygote in equal numbers. On the other hand, in the cross $D \times R$ we have a dominant and a recessive replaced in the next generation by two heterozygotes, while in the cross $H \times H$ half of the offspring return to the homozygous condition. For equilibrium the second type of mating must be twice as frequent as the first, and, if I, J, and K are the means of the distributions of the three phases,

$$4Q^2 e^{\mu J^2/V} = 4PR\, e^{\mu IK/V}.$$

" Since J^2/V and IK/V are small quantities, we shall neglect their squares, and obtain the equation

$$PR - Q^2 = Q^2 \mu \frac{J^2 - IK}{V}. \tag{XIII}$$

If, as before, the two types of gamete are in the ratio $p : q$, the frequencies of the three phases are expressed by the equations

$$\left.\begin{aligned} P &= p^2 + p^2 q^2 \mu \frac{J^2 - IK}{V}, \\ Q &= pq - p^2 q^2 \mu \frac{J^2 - IK}{V}, \\ R &= q^2 + p^2 q^2 \mu \frac{J^2 - IK}{V}. \end{aligned}\right\} \tag{XIV}$$

" It is evident that

$$PI + 2QJ + RK = 0, \tag{XV}$$

and this enables us, whenever necessary, to eliminate J, and to treat only I and K as unknowns. These can only be found when the system of association between different factors has been ascertained. It will be

observed that the changes produced in P, Q, and R are small quantities of the second order: in transforming the quantity

$$p^2q^2\mu\frac{J^2-IK}{V},$$

we may write $-(p^2I+q^2K)$ for $2pqJ$, leading to the form

$$\frac{\mu}{4V}(p^2I-q^2K)^2,$$

which will be found more useful than the other.

" 11. The nine possible combinations of two factors will not now occur in the simple proportions PP', $2PQ'$, etc., as is the case when there is no association: but whatever the nature of the association may be, we shall represent it by introducing new quantities, which by analogy we may expect to be small of the second order, defined so that the frequency of the type

$$DD' \text{ is } PP'(1+f_{11}),$$

that of DH' is $2PQ'(1+f_{12})$,

and that of DR' is $PR'(1+f_{13})$,

and so on."

We now have to study the effect of assortative mating on the joint distribution of pairs of factors since such pairs are not now distributed independently of each other.

Write D, H, R for the phases of one factor with frequencies P, Q, R, and D', H', R' and P', Q', R' for the second factor. The joint frequencies can then be expressed by introducing new quantities $f_{11}, \ldots, f_{33}$ in the manner shown in Table N.

TABLE N

1st factor		2nd factor D' P'	H' Q'	R' R'
D	P	$PP'(1+f_{11})$	$2PQ'(1+f_{12})$	$PR'(1+f_{13})$
H	Q	$2QP'(1+f_{21})$	$4QQ'(1+f_{22})$	$2QR'(1+f_{23})$
R	R	$RP'(1+f_{31})$	$2RQ'(1+f_{32})$	$RR'(1+f_{33})$

(Notice that R and R' are used in two different senses.) Since the sums of the rows and the columns must equal the corresponding row and column frequencies we get (XVI). Since the first three of these equations when multiplied by P, Q, R and added are equal to the second three multiplied by P', Q', R' and added, only five of these equations are independent, and so four of the f's are independent. We take these as f_{11}, f_{13}, f_{31} and f_{33}.

" Formally, we have introduced nine such new unknowns for each pair of factors, but since, for instance, the sum of the above three quantities must be P, we have the six equations

$$\left.\begin{aligned} P'f_{11}+2Q'f_{12}+R'f_{13}=0, \quad & Pf_{11}+2Qf_{21}+Rf_{31}=0,\\ P'f_{21}+2Q'f_{22}+R'f_{23}=0, \quad & Pf_{12}+2Qf_{22}+Rf_{32}=0,\\ P'f_{31}+2Q'f_{32}+R'f_{33}=0, \quad & Pf_{13}+2Qf_{23}+Rf_{33}=0, \end{aligned}\right\} \qquad \text{(XVI)}$$

five of which are independent. The unknowns are thus reduced to four, and we shall use $f_{11}, f_{13}, f_{31}, f_{33}$, since any involving a 2 in the suffix can easily be eliminated.

" We have further

$$\left.\begin{aligned} I &= i+\Sigma(P'i'f_{11}+2Q'j'f_{12}+R'k'f_{13}),\\ J &= j+\Sigma(P'i'f_{21}+2Q'j'f_{22}+R'k'f_{23}),\\ K &= k+\Sigma(P'i'f_{31}+2Q'j'f_{22}+R'k'f_{33}), \end{aligned}\right\} \qquad \text{(XVII)}$$

in which the summation is extended over all the factors except that one to which i, j, k refer. Since we are assuming the factors to be very numerous, after substituting their values for the f's we may without error extend the summation over all the factors. The variance defined as the mean square deviation may be evaluated in terms of the f's

$$V = \Sigma(Pi^2 + 2Qj^2 + Rk^2) + 2\Sigma\{PP'(1+f_{11})\,ii' + 8 \text{ other terms}\},$$

which reduces to

$$\Sigma(Pi^2 + 2Qj^2 + Rk^2) + 2\Sigma\{PP'ii'f_{11} + 8 \text{ other terms}\},$$

so that

$$V = \Sigma(PiI + 2QjJ + RkK).'' \qquad \text{(XVIII)}$$

We are assuming no epistasis, but the non-randomness of mating makes the average value of individuals which are D, H and R for some particular locus not equal to i, j, k, which are the values they would have if the genes at the other loci were fixed. Thus if there are just two loci the average value of individuals which are D for the first locus is got by averaging the deviations of DD', DH', DR', and so is

$$P^{-1}\{(i+i')\,PP'(1+f_{11}) + (i+j')\,2PQ'(1+f_{12}) + (i+k')\,PR'(1+f_{13})\}$$
$$= P^{-1}\{i(PP'(1+f_{11}) + 2PQ'(1+f_{12}) + PR'(1+f_{13})) + P(i'P' + 2j'Q' + k'R')$$
$$+ P(i'P'f_{11} + 2j'Q'f_{12} + k'R'f_{13})\}.$$

By (XVI), the definition of i', j', k', and of f_{11}, f_{12}, f_{13} this is equal to $i + (i'P'f_{11} + 2j'Q'f_{12} + k'R'f_{13})$, and summing over all other factors we obtain (XVII). Notice that we then have

$$PI + 2QJ + RK = 0.$$

Suppose that the biometric measurement can be written as the sum, ΣX_i, of a large number of factors. By definition the mean value of each X_i is zero and the variance is

$$E(\Sigma X_i)^2 = \Sigma EX_i^2 + \sum_{i \neq j} E(X_i X_j),$$

and inserting the above values we get

$$\Sigma(Pi^2 + 2Qj^2 + Rk^2) + 2\Sigma\{PP'(1+f_{11})\,ii' + 2Q'P(1+f_{12})\,ij' + PR'(1+f_{13})\,ik' + 2QP'(1+f_{21})\,ji'$$
$$+ 4QQ'(1+f_{22})\,jj' + 2QR'(1+f_{23})\,jk' + RP'(1+f_{31})\,ki' + 2RQ'(1+f_{32})\,kj' + RR'(1+f_{33})\,kk'\}.$$

The second summation is taken over all distinct pairs of factors. The terms within the second summation not involving f's add to zero, and using (XVII) we obtain (XVIII).

" 12. We can only advance beyond these purely formal relations to an actual evaluation of our unknowns by considering the equilibrium of the different phase combinations. There are forty-five possible matings of the nine types, but since we need only consider the equilibrium of the four homozygous conditions, we need only pick out the terms, ten in each case, which give rise to them. The method will be exactly the same as we used for a single factor. Thus the matings $DD' \times DD'$ have the frequency

$$PP'.PP'.(1+f_{11})(1+f_{11})\exp\{\mu(I+I')^2/V\},$$

which for our purpose is equal to

$$P^2P'^2[1 + 2f_{11} + (\mu/V)(I+I')^2].$$

The number of possible pairs of phases is $9 + \frac{1}{2}(9)(8) = 45$, but we only need to consider the four homozygous types. Then a mating of type $DD' \times DD'$ will have a relative frequency

$$\{PP'(1+f_{11})\}^2 \exp\{\mu(I+I')^2/V\}$$

which is approximately

$$(PP')^2\{1 + 2f_{11} + (\mu/V)(I+I')^2\}.$$

We consider all the matings which give rise to the four homozygous types and it is sufficient, by symmetry, to consider the terms which give rise to DD'. For single factors the only matings

which give D are $D \times D$, $D \times H$ and $H \times H$. Thus the ten relevant matings with their relative frequencies and the proportion of DD' in their offspring are given in Table O.

TABLE O

Mating	Frequency	Probability of DD'
$DD' \times DD'$	$(PP')^2\{1+2f_{11}+(\mu/V)(I+I')^2\}$	1
$DD' \times DH'$	$2P^2P'Q'\{1+f_{11}+f_{12}+(\mu/V)(I+I')(I+J')\}$	$\frac{1}{2}$
$DH' \times DH'$	$4P^2Q'^2\{1+2f_{12}+(\mu/V)(I+J')^2\}$	$\frac{1}{4}$
$DD' \times HD'$	$2PQP'^2\{1+f_{11}+f_{21}+(\mu/V)(I+I')(J+I')\}$	$\frac{1}{2}$
$DD' \times HH'$	$4PQP'Q'\{1+f_{11}+f_{22}+(\mu/V)(I+I')(J+J')\}$	$\frac{1}{4}$
$DH \times HH'$	$8PQQ'^2\{1+f_{12}+f_{22}+(\mu/V)(I+J')(J+J')\}$	$\frac{1}{8}$
$HD' \times HD'$	$4Q^2P'^2\{1+2f_{21}+(\mu/V)(J+I')^2\}$	$\frac{1}{4}$
$HD' \times HH'$	$8Q^2P'Q'\{1+f_{21}+f_{22}+(\mu/V)(J+I')(J+J')\}$	$\frac{1}{8}$
$HH' \times HH'$	$16Q^2Q'^2\{1+2f_{22}+(\mu/V)(J+J')^2\}$	$\frac{1}{16}$
$DH' \times HD'$	$4PQP'Q'\{1+f_{12}+f_{21}+(\mu/V)(I+J')(J+I')\}$	$\frac{1}{4}$

In cases where the pairs of mating individuals are different the above frequencies must be multiplied by two. Adding all together we obtain the left-hand side of equation (XIX). The fact that these together equal the right-hand side expresses the condition that the frequency of DD' does not change from generation to generation.

" Collecting now all the matings which yield DD', we have for equilibrium

$$\begin{aligned}
&P^2P'^2[1+2f_{11}+(\mu/V)(I+I')^2]+2P^2P'Q'[1+f_{11}+f_{12}+(\mu/V)(I+I')(I+J')]\\
&\quad+2PQP'^2[1+f_{11}+f_{21}+(\mu/V)(I+I')(J+I')]+2PQP'Q'[1+f_{11}+f_{22}+(\mu/V)(I+I')(J+J')]\\
&\quad+2PQP'Q'[1+f_{12}+f_{21}+(\mu/V)(I+J')(J+I')]+P^2Q'^2[1+2f_{12}+(\mu/V)(I+J')^2]\\
&\quad+Q^2P'^2[1+2f_{21}+(\mu/V)(J+I')^2]+2PQQ'^2[1+f_{12}+f_{22}+(\mu/V)(I+J')(J+J')]\\
&\quad+2Q^2P'Q'[1+f_{21}+f_{22}+(\mu/V)(J+I')(J+J)']+Q^2Q'^2[1+2f_{22}+(\mu/V)(J+J')^2]\\
&\qquad\qquad=PP'(1+f_{11}) \qquad\qquad \text{(XIX)}
\end{aligned}$$

" Now since

$$(P+Q)^2(P'+Q')^2-PP'(P+2Q+R)(P'+2Q'+R')=(Q^2-PR)P'+(Q'^2-P'R')P+(Q^2-PR)(Q'^2-P'R')$$

the terms involving only P and Q, reduce (XIII) to the second order of small quantities."

Consider all the terms on the left-hand side of (XIX) which do not involve f's or μ. These sum to

$$(P+Q)^2(P'+Q')^2.$$

The equation immediately following (XIX) is an algebraic identity. If quantities such as $IV^{-\frac{1}{2}}$ are regarded as being of the first order of smallness, Q^2-PR, and $Q'^2-P'R'$ are of the second order of smallness and we can neglect $(Q^2-PR)(Q'^2-P'R')$. Hence the difference between the sums of terms on the left- and right-hand sides not involving f's or μ is equal to

$$(Q^2-PR)P'+(Q'^2-P'R')P=-P'Q^2(\mu/V)(J^2-IK)-PQ'^2(\mu/V)(J'^2-I'K')$$

by using (XIII), the error being of the fourth order. There is a misprint in the paper, it being (XIX) which is reduced and not (XIII). Fisher probably means 'by (XIII)'.

Next we pick out of (XIX) the terms involving μ and these sum identically to

$$(\mu/V)\{(P'+Q')(PI+QJ)+(P+Q)(P'I'+Q'J')\}^2.$$

From this we eliminate J and J' by using the identities

$$PI+2QJ+RK=0, \quad P'I'+2Q'J'+R'K'=0.$$

We then obtain

$$(\mu/4V)\{(P'+Q')(PI-RK)+(P+Q)(P'I'-R'K')\}^2=(\mu/4V)\{p'(PI-RK)+p(P'I'-R'K')\}^2,$$

on writing $p'=P'+Q'$, $p=P+Q$. Expanding the square and subtracting the previously obtained term, we get

$$(\mu/2V)pp'(PI-RK)(P'I'-R'K').$$

Next consider the terms on the left-hand side involving f's. Adding, and using p, p', we get

$$2PP'pp'f_{11}+2PQ'pp'f_{12}+2QP'pp'f_{21}+2QQ'pp'f_{22}.$$

We get rid of the suffix 2 by using

$$2Q'f_{12}=-P'f_{11}-R'f_{13},$$
$$2Qf_{21}=-Pf_{11}-Rf_{31},$$
$$4QQ'f_{22}=PP'f_{11}+PR'f_{13}+RP'f_{31}+RR'f_{33},$$

and we obtain

$$\tfrac{1}{2}pp'\{PP'f_{11}-PR'f_{13}-P'Rf_{31}+RR'f_{33}\}.$$

Adding this to the term in μ and equating to the right-hand side we obtain (XIXa). Writing down the three other equations, and adding and subtracting we get (XX) on using $p+q=p'+q'=1$. Substituting back in (XIXa), and putting $P=p^2$, $P'=p^2$ which we can do to the degree of approximation to which we are working, we get the four equations (XXI) which give the f's explicitly.

" $-(\mu/V)[P'Q^2(J^2-IK)+PQ'^2(J'^2-I'K')]=-(\mu/4V)[p'^2(IP-KR)^2+p^2(I'P'-K'R')^2]$.

Also collecting the terms in I and J, we find

$$(\mu/V)[(P'+Q')(IP+JQ)+(P+Q)(I'P'+J'Q')]^2,$$

which yields on eliminating J, $\quad (\mu/4V)[p'(IP-KR)+p(I'P'-K'R')]^2,$

while the result of collecting and transforming the terms in f is

$$\tfrac{1}{2}pp'[PP'f_{11}-PR'f_{13}-P'Rf_{31}+RR'f_{33}].$$

Hence, if the frequency of the type DD' is unchanged

$$(\mu/2V)pp'(IP-KR)(I'P'-K'R')+\tfrac{1}{2}pp'[PP'f_{11}-PR'f_{13}-P'Rf_{31}+RR'f_{33}]=PP'f_{11}. \qquad \text{(XIX}a\text{)}$$

" Now the corresponding equations for the types DR'. RD', $R'D'$ may be obtained simply by substituting K for I. R for P, and vice versa, as required; and each such change merely reverses the sign of the left-hand side, substituting q or q' for p or p' as a factor.

Combining the four equations

$$(\mu/2V)(IP-KR)(I'P'-K'R')=\tfrac{1}{2}[PP'f_{11}-PR'f_{13}-RP'f_{31}+RR'f_{33}] \qquad \text{(XX)}$$

so that the set of four equations

$$(\mu/V)(IP-KR)(I'P'-K'R')=pp'f_{11}=-pq'f_{13}=-qp'f_{31}=qq'f_{33} \qquad \text{(XXI)}$$

gives the whole of the conditions of equilibrium.

" 13. Substituting now in (XVII), which we may rewrite,

$$I=i+\Sigma[P'(i'-j')f_{11}-R'(j'-k')f_{13}],$$
$$K=k+\Sigma[P'(i'-j')f_{31}-R'(j'-k')f_{33}],$$

we have

$$IP-KR=iP-kR+\Sigma(\mu/V)(IP-KR)(I'P'-K'R')[p'(i'-j')+q'(j'-k')]=iP-kR+A(IP-KR),$$

where
$$A(1-A) = (\mu/V)\,\Sigma(i'P'-k'R')\,[p'(i'-j')+q'(j'-k')]$$
$$= (\mu/V)\,\Sigma\beta^2, \text{ since } \beta^2 = \frac{(iP-kR)^2}{2Q}$$

or
$$A(1-A) = \mu(\tau^2/V). \tag{XXII}$$

Using (XVI) we convert (XVII) into
$$I = i+\Sigma\{P'(i'-j')f_{11}-R'(j'-k')f_{13}\},$$
$$K \equiv k+\Sigma\{P'(i'-j')f_{31}-R'(j'-k')f_{33}\}.$$

Here the summation is taken over all loci other than the particular one under consideration. Multiplying by P and R, and subtracting we obtain
$$IP-KR = iP-kR+A(IP-KR),$$

where
$$A = \Sigma(\mu/V)\,(I'P'-K'R')\,\{p'(i'-j')+q'(j'-k')\}.$$

In this form the result is not useful since I' and K', which refer the loci over which the summation is taken, occur on the right-hand side. We therefore apply the same formula as above to each of these loci to obtain
$$I'P'-K'R' = i'P'-K'R'+A(I'P'-K'R'),$$

because the summation can be taken over all loci, the contribution of any particular one being negligible. Then $(1-A)\,(I'P'-K'R') = i'P'-k'R'$, and substituting again we get
$$A = \Sigma\frac{\mu}{V}\frac{i'P'-k'R'}{(1-A)}\{p'(i'-j')+q'(j'-k')\},$$

so that
$$A(1-A) = \Sigma(\mu/V)\,(i'P'-k'R')\,\{p'(i'-j')+q'(j'-k')\}, \tag{XXIIa}$$

and each term in the sum now refers to a single locus. We can therefore drop the dashes. To the degree of approximation required we can put $P = p^2$, $R = q^2$ and
$$p(i-j)+q(j-k) = pi-qk+(p-q)\,(1/2pq)\,(p^2i+q^2k)$$
$$= (1/2pq)\,(p^2i-q^2k),$$

so that finally
$$A(1-A) = \frac{\mu}{V}\Sigma\frac{(p^2i-q^2k)^2}{2pq}$$
$$= \frac{\mu}{V}\Sigma\beta^2 = \frac{\mu\tau^2}{V}. \tag{XXIIb}$$

" It would seem that there is an ambiguity in the value of A, so that the same amount of assortative mating would suffice to maintain two different degrees of association: we have, however, not yet ascertained the value of V. Since this also depends upon A, the form of the quadratic is changed, and it will be seen that the ambiguity disappears.

" Supposing A determinate, we may determine the association coefficients f for
$$\left.\begin{aligned} p^2p'^2f_{11} &= \frac{\mu}{(1-A)^2}\frac{(iP-kR)\,(i'P'-k'R')}{V}pp', \\ p^2q'^2f_{13} &= -\frac{\mu}{(1-A)^2}\frac{(iP-kR)\,(i'P'-k'R')}{V}pq'. \end{aligned}\right\} \tag{XXIII}$$

Hence
$$I = i+\frac{\mu}{(1-A)^2}\frac{iP-kR}{pV}\Sigma[p'(i'-j')+q'(j'-k')]\,(i'P'-k'R')$$
$$= i+\frac{\mu}{(1-A)^2}\frac{iP-kR}{p}\frac{\tau^2}{V},$$

and so
$$I = i + \frac{A}{1-A}\frac{iP-kR}{p}.$$

Similarly
$$K = k - \frac{A}{1-A}\frac{iP-kR}{q}, \qquad \text{(XXIV)}$$

and
$$J = j - \frac{A}{1-A}\frac{p-q}{2pq}(iP-kR).$$

We have
$$\frac{i'P'-k'R'}{I'P'-K'R'} = I - A,$$

and substituting this in (XXI) and multiplying by pp' we get (XXIII) from which (XXIV) follows by simple substitution using (XXIIa) and (XXIIb).

" So that the sense in which the mean value of the heterozygote is changed by assortative mating depends only on whether p or q is greater. In spite of perfect dominance, the mean value of the heterozygote will be different from that of the dominant phase.

" The value of the variance deduced from the expression

$$V = \Sigma(PiI + 2QjJ + RkK)$$

reduces to a similar form. For evidently

$$V = \Sigma\alpha^2 + \frac{A}{1-A}\Sigma(iP-kR)\,[p(i-j)+q(j-k)].$$

Hence
$$V = \sigma^2 + \frac{A}{1-A}\tau^2. \qquad \text{(XXV)}$$

Therefore the equation for A finally takes the form

$$\mu\tau^2 = VA(1-A) = A(1-A)\,\sigma^2 + A^2\tau^2,$$

and may be otherwise written
$$A^2\epsilon^2 - A\sigma^2 + \mu\tau^2 = 0. \qquad \text{(XXVI)"}$$

Here $\epsilon^2 = \sigma^2 - \tau^2$ as usual. When $A = 0$ the left-hand side is $\mu\tau^2 > 0$. When $A = \mu$ it becomes $\mu(\mu-1)\,(\sigma^2-\tau^2)$ which is negative and when $A = 1$ it is still negative, whilst when A is large it is again positive. Thus the quadratic must have two roots, one in the interval $(0, \mu)$ and the other greater than unity. A cannot be greater than unity because the right-hand side of (XXIIb) is positive.

" Now, since the left-hand side is negative when $A = 1$, there can be only one root less than unity. Since, moreover,

$$(\mu - A^2)\,\tau^2 = (A - A^2)\,\sigma^2 \qquad \text{(XXVI}a\text{)}$$

it is evident that this root is less than μ, and approaches that value in the limiting case when there is no dominance.

" A third form of this equation is of importance, for

$$\frac{A}{\mu} = \frac{\tau^2}{\sigma^2 - A\epsilon^2} = \frac{\tau^2 + [A/(1-A)]\tau^2}{\sigma^2 + [A/(1-A)]\tau^2} \qquad \text{(XXVI}b\text{)}$$

which is the ratio of the variance without and with the deviations due to dominance.

" 14. *Multiple Allelomorphism.* The possibility that each factor contains more than two allelomorphs makes it necessary to extend our analysis to cover the inheritance of features influenced by such polymorphic factors. In doing this we abandon the strictly Mendelian mode of inheritance, and treat of Galton's 'particulate inheritance' in almost its full generality. Since, however, well-authenticated cases of multiple allelomorphism have been brought to light by the Mendelian method of research, this generalised conception of inheritance may well be treated as an extension of the classical Mendelism, which we have so far investigated.

" If a factor have a large number, n, of allelomorphs, there will be n homozygous phases, each of which is associated with a certain deviation of the measurement under consideration from its mean value. These deviations will be written $i_1, i_2, \ldots, i_n$, and the deviations of the heterozygous phases, of which there are $\frac{1}{2}n(n-1)$, will be written j_{12}, j_{13}, j_{23}, and so on. Let the n kinds of gametes exist with frequencies proportional to p, q, r, s, and so on, then when the mating is random the homozygous phases must occur with frequencies proportional to $p^2, q^2, r^2, \ldots$, and the heterozygous phases to $2pq, 2pr, 2qr, \ldots$.

" Hence, our measurements being from the mean,

$$p^2 i_1 + q^2 i_2 + r^2 i_3 + \ldots + 2pq j_{12} + 2pr j_{13} + \ldots = 0. \qquad \text{(XII*)}$$

" As before, we define α^2 by the equation

$$p^2 i_1^2 + q^2 i_2^2 + r^2 i_3^2 + \ldots + 2pq j_{12}^2 + 2pr j_{13}^2 + \ldots = \alpha^2 \qquad \text{(I*)}$$

and choosing $l, m, n, \ldots$, so that

$$p^2(2l - i_1)^2 + q^2(2m - i_2)^2 + \ldots + 2pq(l + m - j_{12})^2 + 2pr(l + n - j_{13})^2 + \ldots$$

is a minimum, we define β^2 by

$$4l^2p^2 + 4m^2q^2 + \ldots + 2pq(l+m)^2 + 2pr(l+n)^2 \ldots = \beta^2,$$

the condition being fulfilled if

$$l = pi_1 + qj_{12} + rj_{13} + \ldots,$$

$$m = pj_{12} + qi_2 + rj_{23} + \ldots,$$

and so on.

" Now

$$\beta^2 = S(4l^2p^2) + S(2pq\{l+m\}^2),$$

$$= S(2p(1+p)\,l^2) + S(4pqlm),$$

and since

$$pl + qm + rn + \ldots = 0,$$

$$\beta^2 = S(2pl^2),$$

which may now be written as a quadratic in i and j, represented by the typical terms

$$2p^3 i_1^2 + 4p^2 q i_1 j_{12} + 2pq(p+q) j_{12}^2 + 4pqr j_{12} j_{13}.$$

We assume there are n alleles $A_1, \ldots, A_n$ with frequencies $p, q, r, \ldots$ respectively. The n homozygotes are

$$A_1A_1, \ldots, A_nA_n,$$

with values

$$i_1, \ldots, i_n,$$

and there are $\frac{1}{2}n(n-1)$ heterozygotes $A_1A_2, A_1A_3, \ldots$, whose values are $j_{12}, j_{13}, \ldots$.

Put

$$S_1 = p^2(2l - i_1)^2 + \ldots + 2pq(l + m - j_{12})^2 + \ldots$$

where $l, m, n, \ldots$ are to be chosen by least squares to give the linear additive contribution to the variance. (Fisher uses S without a suffix for summation.)

The minimization equations are typified by

$$0 = \frac{1}{2}\frac{\partial S_1}{\partial l} = p^2(2l - i_1) + pq(l + m - j_{12}) + pr(l + n - j_{13}) + \ldots$$

$$= p\{l(p+1) - pi_1 - qj_{12} - rj_{13} - \ldots + qm + rn + \ldots\}$$

and since $p \neq 0$,

$$l - pi_1 - qj_{12} - rj_{13} - \ldots + (pl + qm + rn + \ldots) = 0.$$

Multiplying this equation by p, the corresponding equation by $q, r, \ldots$, and adding we get

$$(pl + qm + \ldots) + (p + q + \ldots)(pl + qm + \ldots) = 0$$

because we have defined $i_1, i_2, \ldots$ and $j_{12}, \ldots$ so that the population mean

$$p^2 i_1 + q^2 i_2 + \ldots + 2pq j_{12} + \ldots$$

is zero. Hence

$$pl + qm + \ldots = 0,$$

and

$$l = pi_1 + qj_{12} + rj_{13} + \ldots.$$

The linear component of variance is then

$$\beta^2 = 4\{l^2p^2 + m^2q^2 + \ldots\} + 2\{pq(l+m)^2 + \ldots\}.$$

But $pl + qm + \ldots = 0$, and therefore

$$p^2l^2 + q^2m^2 + \ldots + 2pqlm + \ldots = 0.$$

Taking twice this from β^2 we get

$$\begin{aligned}\beta^2 &= (2pl^2 + 2qm^2 + \ldots)(p + q + \ldots)\\ &= 2pl^2 + 2qm^2 + \ldots,\end{aligned}$$

and inserting the values of l, m, ... we get

$$\begin{aligned}\beta^2 &= 2p(pi_1 + qj_{12} + \ldots)^2 + 2q(pj_{12} + qi_2 + rj_{32} + \ldots)^2 + \ldots\\ &= 2(p^3i_1^2 + q^3i_2^2 + \ldots) + 4(p^2qi_1\,j_{12} + pq^2i_2\,j_{12} + \ldots) + 2(pq^2j_{12}^2 + pr^2j_{13}^2 + \ldots + p^2qj_{12}^2 + \ldots)\\ &\quad + 4(pqrj_{12}\,j_{13} + \ldots)\end{aligned}$$

of which the typical term is that given by Fisher.

"Now we can construct an association table for parent and child as in Article 6, though it is now more complicated, since the j's cannot be eliminated by equation (XII*), and its true representation lies in four dimensions; the quadratic in i and j derived from it is, however, exactly one half of that obtained above, so that the contribution of a single factor to the parental product moment is $\frac{1}{2}\beta^2$. Hence the parental correlation is

$$\frac{1}{2}\frac{\tau^2}{\sigma^2},$$

where τ and σ retain their previous meanings."

The association table between parent and offspring could be written down as a

$$\tfrac{1}{2}n(n+1) \times \tfrac{1}{2}n(n+1)$$

table but we need only to write out the typical terms. Part of these can be obtained from the previous parent–offspring table.

For the parental types we can take A_1A_1 and A_1A_2. The possible offspring types are then typified by A_1A_1, A_1A_2, A_1A_3 and A_2A_3. The resulting table is shown as Table P. The covariance,

TABLE P

Offspring type	Parental type A_1A_1	Parental type A_1A_2
A_1A_1	p^3	p^2q
A_1A_2	p^2q	$pq(p+q)$
A_1A_3	p^2r	pqr
A_2A_3	0	pqr

or, as Fisher calls it, the quadratic expression, is then obtained by summing all terms typified by the above, thus giving

$$p^3i_1^2 + q^3i_2^2 + \ldots + 2p^2qi_1\,j_{12} + \ldots + pq(p+q)j_{12}^2 + \ldots + 2pqrj_{12}\,j_{13} + \ldots = \tfrac{1}{2}\beta^2$$

counting all the terms in their proper multiplicity.

" Moreover, from the fraternal table we may obtain a quadratic expression having for its typical terms

$$\tfrac{1}{4}p^2(1+p)^2 i_1^2+\tfrac{1}{2}p^2q^2 i_1 i_2+p^2q(1+p)\,i_1 j_{12}+p^2qr i_1 j_{13},$$

$$\tfrac{1}{2}pq(1+p+q+2pq)j_{12}^2+pqr(1+2p)j_{12}j_{13}+2pqrsj_{12}j_{34},$$

which, when simplified by removing one quarter of the square of the expression in (XII*), becomes

$$\tfrac{1}{4}p^2(1+2p)\,i_1^2+p^2qi_1 j_{12}+\tfrac{1}{2}pq(1+p+q)\,j_{12}^2+pqrj_{12}j_{13},$$

or, simply, $\tfrac{1}{4}(\alpha^2+\beta^2)$."

The fraternal table is rather more complicated to construct. We start from Table Q which gives the possible offspring from all possible types of mating which are 7 in number.

TABLE Q

Mating	Frequency	Offspring
$A_1A_1\times A_1A_1$	p^4	A_1A_1
$A_1A_1\times A_1A_2$	$4p^3q$	$\tfrac{1}{2}A_1A_1+\tfrac{1}{2}A_1A_2$
$A_1A_1\times A_2A_2$	$2p^2q^2$	A_1A_2
$A_1A_1\times A_2A_3$	$4p^2qr$	$\tfrac{1}{2}A_1A_2+\tfrac{1}{2}A_1A_3$
$A_1A_2\times A_1A_2$	$4p^2q^2$	$\tfrac{1}{4}A_1A_1+\tfrac{1}{2}A_1A_2+\tfrac{1}{4}A_2A_2$
$A_1A_2\times A_1A_3$	$8p^2qr$	$\tfrac{1}{4}A_1A_1+\tfrac{1}{4}A_1A_2+\tfrac{1}{4}A_1A_3+\tfrac{1}{4}A_2A_3$
$A_1A_2\times A_3A_4$	$8pqrs$	$\tfrac{1}{4}A_1A_3+\tfrac{1}{4}A_1A_4+\tfrac{1}{4}A_2A_3+\tfrac{1}{4}A_2A_4$

From this table we can pick out the possible pairs of sibs and their relative frequencies, as given in Table R, one sib corresponding to the columns and one to the rows.

TABLE R

		A_1A_1	A_1A_2
		i_1	j_{12}
A_1A_1	i_1	$\tfrac{1}{4}p^2(1+p)^2$	$p^2q(1+p)$
A_1A_2	j_{12}	$p^2q(1+p)$	$\tfrac{1}{2}pq(1+p+q+2pq)$
A_1A_3	j_{13}	$p^2r(1+p)$	$\tfrac{1}{2}pqr(1+2p)$
A_2A_3	j_{23}	p^2qr	$\tfrac{1}{2}pqr(1+2q)$
A_3A_3	i_3	$\tfrac{1}{2}p^2r^2$	pqr^2
A_3A_4	j_{34}	p^2rs	$\tfrac{1}{2}pqrs$

To illustrate how these frequencies are obtained consider the case where both sibs are A_1A_1. This can happen in the first, second, fifth and sixth type of mating and the total frequency is

$$p^4=\tfrac{1}{2}p^3(q+r+\ldots)+\tfrac{1}{4}p^2(q^2+r^2+\ldots)+\tfrac{1}{2}p(qr+qs+\ldots+rs+\ldots)$$
$$=p^4+\tfrac{1}{2}p^3(1-p)+\tfrac{1}{4}p^2(1-p)^2=\tfrac{1}{4}p^2(1+p)^2.$$

(This is more easily obtained by the Li and Sacks method mentioned before.) Adding together all the resulting terms we get

$$\tfrac{1}{4}p^2(1+p)^2i_1^2+\tfrac{1}{4}q^2(1+q)^2i_2^2+\ldots+\tfrac{1}{2}p^2q^2i_1i_2+\tfrac{1}{2}p^2r^2i_1i_3+\ldots+p^2q(1+p)i_1j_{12}+p^2r(1+p)i_1j_{13}+\ldots$$
$$+p^2qri_1j_{23}+p^2qsi_1j_{24}+\ldots+\tfrac{1}{2}pq(1+p+q+2pq)j_{12}^2+\tfrac{1}{2}pr(1+p+r+2pr)j_{13}^2+\ldots$$
$$+pqr(1+2p)j_{12}j_{13}+pqs(1+2p)j_{12}j_{14}+\ldots+2pqrsj_{12}j_{34}+\ldots,$$

thus agreeing with Fisher's sum of typical terms except for his fourth term which should read $p^2qri_1j_{23}$ and not $p^2qri_1j_{13}$.

The square of the expression in (XII*) is

$$\{p^2i_1+q^2i_2+\ldots+2pqj_{12}+2prj_{13}+\ldots\}^2=0,$$

and subtracting $\tfrac{1}{4}$ of this from the above we get $\tfrac{1}{4}(\alpha^2+\beta^2)$ as stated.

" Here, again, the introduction of multiple allelomorphism does not affect the simplicity of our results; the correlation between the dominance deviations of siblings is still exactly $\frac{1}{4}$, and the fraternal correlation is diminished by dominance to exactly one half the extent suffered by the parental correlation. The dominance ratio plays the same part as it did before, although its interpretation is now more complex. The fraternal correlation may be written, as in Article 6,

$$\frac{1}{2\sigma^2}(\tau^2+\tfrac{1}{2}\epsilon^2).$$

" 15. *Homogamy and Multiple Allelomorphism.* The proportions of these different phases which are in equilibrium when mating is assortative must now be determined. As in Article 10, let $I_1, I_2, \ldots$ be the mean deviations of the homozygous phases, and $J_{12}, J_{13}, \ldots$ those of the heterozygous phases. Let the frequency of the first homozygous phase be written as $p^2(1+f_{11})$, and the others in the same way. Then, since p is the frequency of the first kind of gamete,

$$pf_{11}+qf_{12}+rf_{13}+\ldots=0,$$

and

$$pf_{12}+qf_{22}+rf_{23}+\ldots=0,$$

and so on.

" Let

$$pI_1+qJ_{12}+rJ_{13}+\ldots=L,$$

$$pJ_{12}+qI_2+rJ_{23}+\ldots=M,$$

and so on, then $L, M, \ldots$ represent the mean deviations of individuals giving rise to gametes of the different kinds; hence, by Article 9,

$$2pq(1+f_{12}) = 2pq\,e^{\mu/V\,.\,LM},$$

that is,

$$f_{12} = \mu/V\,.\,LM. \qquad \text{(XIV*)}$$

The aim of paragraph 15 is to extend the treatment of assortative mating in paragraphs 9–13 to the case where each locus may have more than two alleles, all loci remaining, as before, unlinked. Since we are concerned with second-degree statistics (variances and covariances) it is sufficient to consider the loci in pairs.

In the stable population with assortative mating $I_1, I_2, \ldots$ and $J_{12}, J_{13}, \ldots$ are taken as the mean values of the deviations from the population mean of the respective homozygotes $A_1A_1, A_2A_2, \ldots$ and the heterozygotes A_1A_2, $A_1A_3, \ldots$, with frequencies $p^2(1+f_{11}), q^2(1+f_{22}), \ldots$ and $pq(1+f_{12})$, $pr(1+f_{13})$, etc. Then the equations such as

$$pf_{11}+qf_{12}+\ldots=0$$

are necessary in order that the gene frequencies amongst all mating pairs should be exactly p, q, etc.

Notice in particular that $f_{11}, f_{12}, \ldots$ are not analogous to the $f_{11}, \ldots$ used in the previous discussion of assortative mating where there are only two alleles at each locus. The f's here refer to a single locus, and when referring to another locus we shall write $f'_{11}, f'_{12}, \ldots$.

The average deviation of the class of individuals which give rise to the gamate A_1 will be

$$(1/2p)\{2p^2I_1+2pqJ_{12}+\ldots\}=L, \qquad \text{(XIV*}a\text{)}$$

to the first approximation, there being further terms involving f's which we can ignore. By the type of argument used before we then have

$$2pq(1+f_{12}) = 2pq\exp\left\{\frac{\mu LM}{V}\right\},$$

and

$$f_{12} = \mu LM/V.$$

The frequencies of A_1A_2 and B_1B_2 are $2pq(1+f_{12})$ and $2p'q'(1+f'_{12})$, and their joint frequency which we now want to find is written as

$$4pqp'q'(1+f'_{12.12})$$

or as

$$4pqp'q'(1+f_{12})(1+f'_{12})(1+f_{12.12}).$$

Note that the absence or presence of a dash on f'_{12} means that f_{12} refers to A_1A_2, and f'_{12} to B_1B_2, whilst on the other hand $f_{12.12}$ and $f'_{12.12}$ both refer to both factors together, the difference being in their definition. Since the f's are all small we expand the products and neglect small terms, obtaining

$$f'_{12.12} = f_{12} + f'_{12} + f_{12.12}.$$

In the absence of any assortative mating the gametic frequency of A_1B_1 would have been pp', but when $\mu \neq 0$ the proportionate increase, using the definition of $f'_{12.12}$, etc., must be

$$pp'f'_{11.11} + pq'f'_{11.12} + pr'f'_{11.13} + \ldots + qp'f'_{12.11} + qq'f'_{12.12} + qr'f'_{12.13} + \ldots + rp'f'_{13.11} + \ldots = F_{11}, \text{ by definition.}$$

Thus the frequency of A_1B_1 in the gametes is $pp'(1+F_{11})$. The mean value of individuals giving rise to this gamete is $L+L'$ by the argument leading to (XIV*a) and so on, so that (XIX*) follows. In this equation the F's are functions of the $f'_{12.12}$, etc., which are known when the frequencies $p, p', \ldots$ are given, and we want to solve for the $f_{12.12}$, etc.

Fisher guesses that the solutions must be

$$f_{12.12} = (\mu/V)(L+M)(L'+M')$$

and similar formulae. Putting these in the equation for F_{11} we get

$$\begin{aligned}
&(\mu/V)\{pp'(2L)(2L') + pq'(2L)(L'+M') + \ldots + qp'(L+M)(2L') \\
&\quad + qq'(L+M)(L'+M') + \ldots + rp'(L+N)(2L') + \ldots\} \\
&= (\mu/V)\{L + pL + qM + \ldots\}\{L' + pL' + qM' + \ldots\} \\
&= (\mu/V)LL',
\end{aligned}$$

since

$$pL + qM + \ldots = 0, \quad pL' + qM' + \ldots = 0.$$

"The association between the phases of two different factors requires for its representation the introduction of association coefficients for each possible pair of phases. Let the homozygous phases of one factor be numbered arbitrarily from 1 to m, and those of the other factor from 1 to n, then, as the phase (12) of the first factor occurs with frequency $2pq(1+f_{12})$, and of the second factor, with frequency $2p'q'(1+f'_{12})$, we shall write the frequency with which these two phases coincide in one individual as $4pqp'q'(1+f'_{12.12})$, or as

$$4pqp'q'(1+f_{12})(1+f'_{12})(1+f_{12.12}),$$

so that

$$f'_{12.12} = f_{12.12} + f_{12} + f'_{12}.$$

"The proportional increase of frequency of the gametic combination (1.1) is

$$pp'f'_{11.11} + pq'f'_{11.12} + pr'f'_{11.13} + \ldots + qp'f'_{12.11} + qq'f'_{12.12} + qr'f'_{12.13} + \ldots,$$

and so on.

"By virtue of the equations connecting the f's of a single factor, this expression, which we shall term F_{11}, has the same value, whether written with dashed or undashed f's.

"Individuals having the constitution (12.12) may be formed by the union either of gametes (1.1) and (2.2), or of gametes (1.2) and (2.1); hence the equations of equilibrium are of the form

$$2f'_{12.12} = F_{11} + F_{22} + (\mu/V)(L+L')(M+M') + F_{12} + F_{21} + (\mu/V)(L+M')(M+L'),$$

but

$$\begin{aligned}
2f_{12.12} &= 2f'_{12.12} - 2f_{12} - 2f'_{12} \\
&= 2f'_{12.12} - (2\mu/V)(LM + L'M'),
\end{aligned}$$

therefore

$$2f_{12.12} = F_{11} + F_{22} + F_{12} + F_{21} + (\mu/V)(L+M)(L'+M'). \qquad \text{(XIX*)}$$

"By analogy with Article 12, the solution

$$f_{12.12} = (\mu/V)(L+M)(L'+M')$$

suggests itself, and on trial it leads to $F_{11} = (\mu/V)LL'$, and is thereby verified."

To obtain L we argue as follows. Write the mean deviation of the homozygote A_1A_1 from the population mean as $I_1 = i_1 + i_1^*$ where i_1 is the deviation due directly to the genotype A_1A_1 if the other genes were segregating independently of the A locus. i_1^* is then the extra deviation produced by the other genes in virtue of the association due to homogamy. The homozygote A_1A_1 has frequency

$$p^2(1+f_{11})$$

and the double homozygote $A_1A_1B_1B_1$ has frequency

$$p^2p'^2(1+f_{11})(1+f'_{11})(1+f_{11,11}).$$

From this it follows that the conditional probability that an individual is B_1B_1, if it is known that it is A_1A_1, is

$$p'^2(1+f'_{11})(1+f_{11,11}),$$

which to a reasonable approximation can be written as

$$p'^2(1+f'_{11}+f_{11,11}).$$

Similarly, the probability that an individual is B_1B_2 when it is known that it is A_1A_1, is approximately

$$2p'q'(1+f'_{12}+f_{11,12}).$$

The total additional contribution of the individuals at the B locus to the measurement of A_1A_1 individuals is therefore

$$p'^2(1+f'_{11}+f_{11.11})\,i'_1+q'^2(1+f'_{22}+f_{11.22})\,i'_2+\ldots+2p'q'(1+f'_{12}+f_{11.12})\,j'_{12}+\ldots.$$

We have already defined the effects i'_r, j'_{rs}, in such a way that the mean effect is zero, i.e. so that

$$p'^2(1+f'_{11})\,i'_1+\ldots+2p'q'(1+f'_{12})\,j'_{12}+\ldots=0.$$

Thus the additional increment is simply

$$p'^2f_{11.11}\,i'_1+\ldots+2p'q'f_{11.12}\,j'_{12}+\ldots.$$

We now consider the sum over all loci other than the A locus and we denote this summation by the symbol Σ. We get

$$I_1 = i_1+i_1^*+\Sigma\{p'^2f_{11.11}\,i'_1+\ldots+2p'q'f_{11.12}\,j'_{12}+\ldots\}$$

and similarly

$$J_{12} = j_{12}+\Sigma\{p'^2f_{12.11}\,i'_1+\ldots+2p'q'f_{12.12}\,j'_{12}+\ldots\},$$

the factor 2 occurring to include two terms of equal value.

Write

$$l = pi_1+qj_{12}+rj_{13}+\ldots.$$

Then

$$\begin{aligned}L &= pI_1+qJ_{12}+\ldots\\ &= l+\Sigma\{p'^2i'_1(f_{11.11}p+f_{12.11}q+\ldots)+\ldots+2p'q'j'_{12}(f_{11.12}p+f_{12.12}q+\ldots)+\ldots\}.\end{aligned}$$

Using the values of the f's which we have found above, and

$$pL+qM+\ldots=0,$$

we get

$$f_{11.11}p+f_{12.11}q+\ldots=(\mu/V)\,L(2L'),$$

and similarly

$$f_{11.12}p+f_{12.12}q+\ldots=(\mu/V)\,L(L'+M')$$

(these results being misprinted in the text). Using these we get finally

$$L = l+(\mu L/V)\,\Sigma\{(2L')\,p'^2i'_1+\ldots+2(L'+M')\,p'q'j'_{12}+\ldots\}.$$

Since $$l' = p'i'_1 + q'j'_{12} + \dots$$

this can be written $$L = l + (\mu L/V)\,\Sigma\{2p'l'L' + 2q'm'M' + \dots\}.$$

Since each individual locus is regarded as contributing a vanishingly small component of the total variance we can now suppose the summation to be taken over all factors at all loci and not merely all those other than the locus being considered. We can then put $L = l + AL$, where

$$A = (\mu/V)\,\Sigma\{2p'l'L' + 2q'm'M' + \dots\}.$$

Since A is independent of the locus being considered, we also have

$$L' = l' + AL',$$

so that $$l' = (1-A)\,L',$$

and then
$$\begin{aligned} A(1-A) &= (\mu/V)\,\Sigma(1-A)\,(2p'l'L' + \dots) \\ &= (\mu/V)\,\Sigma(2p'l'^2 + 2q'm'^2 + \dots) \\ &= (\mu/V)\,\tau^2. \end{aligned} \qquad \text{(XXII*)}$$

In a similar way substituting for the f's in the formulae for I_1 and J_{12}, and then putting $L = l(1-A)^{-1}$, etc. we get

$$\begin{aligned} I_1 &= i_1 + \frac{2\mu}{V(1-A)^2}\{4p'^2ll'i'_1 + \dots + 2p'q'l(l'+m')j_{12} + \dots\} \\ &= i_1 + \frac{2Al}{1-A}, \end{aligned}$$

and
$$\begin{aligned} J_{12} &= j_{12} + \frac{2\mu}{V(1-A)^2}\{2p'^2(l+m)\,l'i'_1 + \dots + 4p'q'(l+m)\,(l'+m')j_{12} + \dots\} \\ &= j_{12} + \frac{A}{1-A}(l+m). \end{aligned}$$

" Hence we may evaluate $L, L', \dots,$ for

$$L = pI_1 + qJ_{12} + rI_{13} + \dots + l = \Sigma\{p'^2i'(pf_{11.11} + qf_{12.11} + \dots) + 2p'q'j'_{12}(pf_{11.12} + qf_{12.12} + \dots) + \dots,$$

but $$pf_{11.11} + qf_{12.11} + \dots = (\mu/V)\,L(L' + M'),$$

therefore
$$\begin{aligned} L &= l + (\mu/V)\,L\Sigma\{p'^2i'(L'+L') + 2p'q'j'_{12}(L'+M') + \dots\}, \\ &= l + (\mu/V)\,L\Sigma(2p'l'L' + 2q'm'M' \dots). \end{aligned}$$

"Let $$L = l + AL,$$

then $$L = l/(1-A),$$

and $$A = (\mu/V)\,\Sigma(2p'l'L' + 2q'm'M' + \dots),$$

therefore
$$\begin{aligned} A(1-A) &= (\mu/V)\,\Sigma(2p'l'^2 + 2q'm'^2 + \dots) \\ &= (\mu/V)\,\Sigma\beta'^2, \end{aligned}$$

therefore $$A(1-A) = (\mu/V)\,\tau^2 \qquad \text{(XXII*)}$$

so that the association constant, A, appearing now in the constant ratio $l : L$, plays exactly the same part in the generalised analysis as it did in the simpler case.

" It may now be easily shown that the mean deviations, I and J, may be calculated from the equations

$$\left.\begin{aligned} I_1 &= i_1 + 2Al/(1-A), \\ \text{and} \quad J_{12} &= j_{12} + [A/(1-A)]\,(l+m), \end{aligned}\right\} \qquad \text{(XXIV*)}$$

and that the variance reduces, as before, to

$$\sigma^2 + [A/(1-A)]\,\tau^2. \qquad \text{(XXV*)}$$

" 16. *Coupling.* In much modern Mendelian work coupling plays an important part, although the results of different investigators do not seem as yet to converge upon any one uniform scheme of coupling. The type found by Morgan in the American Fruit Fly (*Drosophila*) is, however, of peculiar simplicity, and may be found to be the general type of the phenomenon.

" An individual heterozygous in two factors may owe its origin to the union of either of two pairs of gametes either (1.1) × (2.2) or (1.2) × (2.1); when coupling occurs, the gametes given off by such an individual, of all these four types, do not appear in equal numbers, preference being given to the two types from which the individual took its origin. Thus in a typical case these two types might each occur in 28 per cent of the gametes and the other two types in 22 per cent. Coupling of this type is reversible, and occurs with equal intensity whichever of the two pairs are supplied by the grandparents. We may have any intensity from zero, when each type of gamete contributes 25 per cent to complete coupling, when only the two original types of gamete are formed, and the segregation takes place as if only one factor were in action.

The above analysis of polymorphic factors enables us to compare these two extreme cases; for there are 9 phase combinations of a pair of dimorphic factors, or, if we separate the two kinds of double heterozygote, 10, which, apart from inheritance, can be interpreted as the 4 homozygous and the 6 heterozygous phases of a tetramorphic factor. The 4 gametic types of this factor are the 4 gametic combinations (1.1), (1.2), (2.1), (2.2)."

This mapping of a system with two factors at each of two loci on to a system with four factors at a single locus is particularly interesting and can be illustrated as follows.

Suppose that at the first locus the two factors denoted by 1 and 2 in the text are A_1 and A_2, and similarly B_1, B_2 at the second locus. The nine phase combinations are then

$$\begin{array}{lll} A_1A_1B_1B_1 & A_1A_2B_1B_1 & A_2A_2B_1B_1 \\ A_1A_1B_1B_2 & A_1A_2B_1B_2 & A_2A_2B_1B_2 \\ A_1A_1B_2B_2 & A_1A_2B_2B_2 & A_2A_2B_2B_2. \end{array}$$

When linkage ('coupling' is Fisher's term) is considered the double heterozygote $A_1A_2B_1B_2$ really corresponds to two different heterozygotes according as whether A_1 and B_1 are on the same chromosome or on different chromosomes. We shall denote these two types by $(A_1B_1)(A_2B_2)$ and $(A_1B_2)(A_2B_1)$.

Now consider a single locus with four factors C_1, C_2, C_3 and C_4. This results in four homozygotes and six heterozygotes. If we identify the four factors C_1, C_2, C_3 and C_4 with the pairs of factors A_1B_1, A_1B_2, A_2B_1 and A_2B_2 respectively we have the following mapping of the two loci situation on the single locus situation.

$A_1A_1B_1B_1$	C_1C_1	$(A_1B_2)(A_2B_1)$	C_2C_3
$A_1A_1B_1B_2$	C_1C_2	$A_1A_2B_2B_2$	C_2C_4
$A_1A_1B_2B_2$	C_2C_2	$A_2A_2B_1B_1$	C_3C_3
$A_1A_2B_1B_1$	C_1C_3	$A_2A_2B_1B_2$	C_3C_4
$(A_1B_1)(A_2B_2)$	C_1C_4	$A_2A_2B_2B_2$	C_4C_4

Thus to deal with linkage Fisher considers the two possible extreme cases of no linkage and complete linkage when there is assortative mating but, as above, no epistatic effects.

Case I. Here we have two unlinked loci with A_1, A_2 at one, and B_1, B_2 at the other. These have, as usual, gene frequencies p, q, p', q', respectively, and as before the coefficient of assortative mating is μ. The mean deviations associated with A_1A_1, A_1A_2, A_2A_2 are again i, j, k, and i', j', k' for the other locus.

Let L be the mean deviation produced in the population by a gamete carrying A_1, and define M, L', M', similarly. Thus the mean deviations associated with gametes A_1B_1, A_1B_2, A_2B_1, and

A_2B_2 are $L+L'$, $L+M'$ (not $M+M'$ as in Fisher), $M+L'$, and $M+M'$. By the theory given above the frequencies of these four gametes are

$$pp'\{1+(\mu/V)\,LL'\}, \qquad pq'\{1+(\mu/V)\,LM'\},$$
$$qp'\{1+(\mu/V)\,ML'\}, \qquad qq'\{1+(\mu/V)\,MM'\}.$$

These are denoted by Fisher as $\mathbf{p}, \mathbf{q}, \mathbf{r}, \mathbf{s}$.

The frequency of a double homozygote like $A_1A_1B_1B_1$ is, by the previous argument,

$$p^2p'^2(1+f_{11.11}) = p^2p'^2(1+f_{11}+f'_{11}+f_{11.11})$$

approximately, where

$$f_{11} = \mu L^2/V, \quad f'_{11} = \mu L'^2/V, \quad f_{11.11} = 4\mu LL'/V.$$

Thus to this order of approximation the frequency of $A_1A_1B_1B_1$ is

$$p^2p'^2\{1+(\mu/V)\,(L^2+L'^2+4LL')\} = \{pp'(1+(\mu/V)\,LL')\}^2\{1+(\mu/V)\,(L+L')^2\}.$$

Case II. If there is complete linkage the four pairs A_1B_1. A_1B_2, A_2B_1, A_2B_2 each behave like a single gene. We suppose they each have the frequencies given above as $\mathbf{p}, \mathbf{q}, \mathbf{r}, \mathbf{s}$. We also suppose that the deviations produced by these 'genes' are the same as occurred in the previous case so that, for example, a zygote $A_1A_1B_1B_2$ would have the deviation $i+j'$, the genes at any other loci being held fixed.

We must first investigate whether the genotypic frequencies in the second case will be the same as in the first. If there is no assortative mating ($\mu = 0$) this is obviously true since the frequency of $A_1A_1B_1B_1$ in the unlinked system will be $(pp')^2$ which is the square of the frequency, pp', of the 'gene' A_1B_1 in the linked system.

We have also seen that assortative mating changes the frequency of gene combinations at any pair of loci only by a quantity of the first order of smallness. Thus to this degree of approximation the 'genotypic' frequencies should remain the same.

The mean deviation in individuals carrying the gamete A_1B_1 will then be the same in both systems. This is $L+L'$ which Fisher writes as a capital $\mathbf{L}$. The similar result holds for the other gametes.

The variance, V, in the population in the two cases should also be the same.

Then treating A_1B_1, etc., as single genes the frequency of a genotype such as $(A_1B_1)\,(A_1B_1)$ will be, to the first order of approximation,

$$\mathbf{p}^2\{1+(\mu/V)\,\mathbf{L}\}^2 = \{pp'(1+(\mu/V)\,LL')\}^2\{1+(\mu/V)\,(L+L')^2\},$$

which agrees exactly with the result obtained in Case I above. Thus to this degree of approximation, which is as far as Fisher goes in his theory, the two systems of completely unlinked and completely linked genes agree as regards the frequencies of occurrence, the magnitudes of the effects produced by genes and gene-combinations, and the effect of assortative mating.

Fisher does not explicitly prove that the correlation between relatives will be the same in the two cases. To show this it is necessary to show that the values of $\tau^2 = \Sigma\beta^2$ are equal since the correlations will be later expressed in terms of τ^2, V, μ, and A (A being given by equation (XXII)).

To prove this we return to the original definition of β^2. To simplify the notation denote the mean deviations i, j, k produced by A_1A_1, A_1A_2, A_2A_2 by $j_{11}, j_{12} = j_{21}, j_{22}$ (notice the change in notation from Fisher's use of these symbols). We shall also write the gene frequencies p, q of A_1, A_2 as

p_1, p_2. As before we proceed by fitting 'representative values' for which we shall abandon Fisher's notation $c+b$, c, $c-b$, and write instead

$$\text{Representative value for } A_r A_s = x_r + x_s,$$

where $r, s = 1, 2$. These values are to be found by least squares. Neglecting the small changes in frequency due to assortative mating we have to minimize the sum

$$S_1 = \Sigma p_r p_s (j_{rs} - x_r - x_s)^2,$$

which is equal to $$p^2(i-c-b)^2 + 2pq(j-c)^2 + q^2(k-c+b)^2$$

in Fisher's notation. The conditions for a minimum are that

$$\frac{1}{2}\frac{\partial S_1}{\partial p_r} = \sum_r p_s (j_{rs} - x_r - x_s) = 0.$$

β^2 is then defined as the variance of the representative values so that

$$\begin{aligned}\beta^2 &= p^2 i^2 + 2pqj^2 + q^2k^2 - (p^2 i + 2pqj + q^2 k)^2 \\ &= \Sigma p_r p_s (x_r + x_s)^2 - \{\Sigma p_r p_s (x_r + x_s)\}^2 \\ &= 2(\Sigma p_r x_r^2 - M^2),\end{aligned}$$

where $$M = \Sigma p_r x_r.$$

The same argument applies if we have multiple allelomorphs $A_r (r = 1, \ldots, k)$, and a similar definition applies to the alleles at the second locus for which the representative values, $x'_r + x'_s$, are the solutions of

$$\sum_r p'_s (j'_{rs} - x'_r - x'_s) = 0.$$

Now consider the system with complete linkage so that the 'alleles' are $(A_r B_s)$. Since there is no epistacy, the mean deviation produced by $A_r A_s B_t B_u$ is $j_{rs} + j_{tu}$, all other genes being fixed. If we neglect the small deviations in frequency produced by assortative mating we can find a 'representative value', x_{rs}, for the 'gene' $(A_r B_s)$ by minimizing the sum

$$\Sigma p_r p_s p'_t p'_u (j_{rs} + j_{tu} - x_{rt} - x_{su})^2.$$

The conditions for this are, by differentiating,

$$2\sum_{rt} p_s p_u (j_{rs} + j_{tu} - x_{rt} - x_{su}) = 0.$$

The solution of these equations is simply $x_{rt} = x_r + x'_t$ as can be verified by substituting these values and using the previous equations for x_r, x'_t.

The new value of β^2 for the system with complete linkage is

$$\begin{aligned}\beta''^2 &= 2\Sigma p_r p'_t (x_r + x'_t)^2 - 2\{\Sigma p_r p'_t (x_r + x'_t)\}^2 \\ &= 2\{\Sigma p'_t \Sigma p_r x_r^2 + 2\Sigma p_r x_r \Sigma p'_t x'_t + \Sigma p_r \Sigma p'_t x'^2_t - (\Sigma p'_t \Sigma p_r x_r + \Sigma p_r \Sigma p'_t x'_t)^2\} \\ &= 2\{\Sigma p_r x_r^2 + 2MM' + \Sigma p'_t x'^2_t - (M + M')^2\} \\ &= 2\{\Sigma p_r x_r^2 - M^2 + \Sigma p'_t x'^2_t - M'^2\} \\ &= \beta^2 + \beta'^2.\end{aligned}$$

Thus in the sum $\tau^2 = \Sigma\beta^2$ the two terms β^2 and β'^2 which occur in the system with unlinked genes are replaced by a single term β''^2 in the system with complete linkage, but by the above

equation the value of r^2 is unchanged. Thus, as will be shown later, the correlations between relatives are unaffected.

Fisher does not consider what happens with values of the recombination fraction lying between 0 and $\frac{1}{2}$, and he seems to imply that because there is no important difference between the extreme cases of no linkage and complete linkage it is highly probable that the same results will be obtained for such intermediate values.

However, there are serious gaps in the argument to be filled before this is demonstrated. It might be thought that a population in which the recombination value was inside the interval $(0, \frac{1}{2})$ could be regarded as a mixture of two populations in one of which linkage is absent and in the other in which it is complete. Simple calculations show that this is not correct.

When linkage is complete the gene combinations, $(A_r B_s)$, can be regarded as single genes and there are no restrictions on the frequencies which can be assigned to them. In particular we can suppose, as above, that $(A_r B_s)$ has frequency $p_r p'_s(1+F_{rs})$. But if linkage is not complete the frequencies of gene combinations are determined by the properties of the system and can no longer be chosen at will. It is therefore of interest to show that the frequency of $A_r B_s$ can still be taken as $p_r p'_s$ in a stable population.

When the mating is not assortative, and $F_{rs} = 0$, this is well known. It can be proved for assortative mating in the following way. Suppose first that the two loci are unlinked. Then a double heterozygote, $A_1 A_2 B_1 B_2$, can arise in two ways. Either $A_1 B_1$ comes from one parent and $A_2 B_2$ from the other (call this 'coupling'), or $A_1 B_2$ comes from one, and $A_2 B_1$ from the other ('repulsion'). The frequencies of $A_1 B_1$ and $A_2 B_2$ are

$$pp'(1+F_{11}) = pp'\{1+(\mu LL'/V)\},$$

and

$$qq'\{1+(\mu MM'/V)\}.$$

The average deviation of individuals giving rise to $A_1 B_1$ is $L+L'$, and that of individuals giving rise $A_2 B_2$ is $M+M'$. If there was no assortative mating the probability of such a pair of gametes would be the product of their frequencies. However, with assortative mating this product has to be multiplied by

$$\exp[(\mu/V)(L+L')(M+M')],$$

which can be approximated by

$$1+(\mu/V)(L+L')(M+M').$$

Thus with assortative mating the total probability of such a pair of gametes is

$$pp'qq'\{1+(\mu/V)[LL'+MM'+(L+L')(M+M')]\}$$
$$= pp'qq'\{1+(\mu/V)(LM+L'M'+(L+M)(L'+M'))\}.$$

By symmetry we get the same probability of a union between $A_1 B_2$ and $A_2 B_1$ so that coupling and repulsion are equally frequent.

Suppose that we have a stable population in which there is no linkage, and instantaneously linkage is introduced with recombination fraction c where $0 < c < \frac{1}{2}$. In the immediately following generation the only effect which could occur would be a change in the proportion of offspring of double heterozygotes. From a double heterozygote in coupling we get gametes in the proportion

$$\tfrac{1}{2}(1-c)A_1 B_1 + \tfrac{1}{2}cA_1 B_2 + \tfrac{1}{2}cA_2 B_1 + \tfrac{1}{2}(1-c)A_2 B_2,$$

and from one in repulsion we get:

$$\tfrac{1}{2}cA_1 B_1 + \tfrac{1}{2}(1-c)A_1 B_2 + \tfrac{1}{2}(1-c)A_2 B_1 + \tfrac{1}{2}cA_2 B_2.$$

Since coupling and repulsion heterozygotes have the same phenotype, they have the same probability of mating with any particular genotype, and since coupling and repulsion are equally frequent, the gametes produced by all heterozygotes will have frequencies which are the averages of the above frequencies, i.e. $\frac{1}{4}A_1B_1+\frac{1}{4}A_1B_2+\frac{1}{4}A_2B_1+\frac{1}{4}A_2B_2$,

which is just what happens if $c=\frac{1}{2}$, i.e. when there is no linkage.

Since the introduction of linkage has not changed the frequencies in the next generation the population remains stable in all further generations.

The same argument can be used to show that the parent–offspring correlation is independent of the recombination fraction. It does not show at once that the same is true for sib–sib and more distant relationships but this is plausible. Fisher does not discuss these more complicated cases in his paper and we do not pursue the matter further.

"The mean deviations associated with these 4 gametic types are $L+L'$, $M+M'$, ..., and we therefore write

$$\mathbf{L}=L+L',\quad \mathbf{M}=L+M',\quad \mathbf{N}=M+L',\quad \mathbf{O}=M+M'.$$

"Further, if these gametic types occur with frequency,

$$\mathbf{p}=pp'\{1+(\mu/V)\,LL'\}\quad \mathbf{q}=pq'\{1+(\mu/V)\,LM'\}$$
$$\mathbf{r}=qp'\{1+(\mu/V)\,ML'\}\quad \mathbf{s}=qq'\{1+(\mu/V)\,MM'\},$$

it is clear that the frequencies with which the homozygous phases occur, such as

$$p^2p'^2(1+f'_{11.11})=p^2p'^2\{1+(\mu/V)\,(L^2+L'^2+4LL')\},$$
$$\mathbf{p}^2\{1+(\mu/V)\,(L+L')^2\}=\mathbf{p}^2(1+(\mu/V)\,\mathbf{L}^2),$$

are exactly those produced, if there really were a single tetramorphic factor.

"In the same way the phases heterozygous in one factor also agree, for

$$2p^2p'q'(1+f'_{11.12})=2p^2p'q'\{1+(\mu/V)\,L^2+L'M'+2L(L'+M'))\}$$
$$=2\mathbf{pq}\{1+(\mu/V)\,(L+L')\,(L+M')\}=2\mathbf{pq}\{1+(\mu/V)\,\mathbf{LM}\}.$$

"Finally, taking half the double heterozygotes,

$$2pqp'q'(1+f'_{12.12})=2pqp'q'\{1+(\mu/V)\,[LM+L'M'+(L+M)\,(L'+M')]\}$$
$$2\mathbf{ps}\{1+(\mu/V)\,(L+L')\,(M+M')\}=2\mathbf{ps}\{1+(\mu/V)\,\mathbf{LO}\},$$

or, equally,
$$2\mathbf{qr}\{1+(\mu/V)\,(L+M')\,(M+L')\}=2\mathbf{qr}\{1+(\mu/V)\,\mathbf{MN}\}.$$

"From this is appears that a pair of factors is analytically replaceable by a single factor if the phase frequencies be chosen rightly: but the only difference in the inheritance in these two systems is that in the one case there is no coupling, and in the other coupling is complete. It would appear, therefore, that coupling is without influence upon the statistical properties of the population."

Fisher now considers the correlations between individuals in a population in which there is assortative mating and environmental effects. To do this he uses regression theory. Suppose all measurements are taken from the mean of the population. Let x be the measurement in one individual and X in another. The correlation between x and X is then

$$\rho=\operatorname{cov}(x,X)\{\operatorname{var}(x)\operatorname{var}(X)\}^{-\frac{1}{2}}$$
$$=\operatorname{cov}(x,X)/V.$$

We suppose so many factors are acting that the joint distribution of x and X is bivariate normal so that the regression lines are straight. Then the expected value of X for any given x is

$$E(X\,|\,\text{given}\,x)=\beta x=\rho x\quad\text{and}\quad \rho=x^{-1}E(X\,|\,\text{given}\,x).$$

Fisher tacitly supposes that the effects of environment can be represented by an addition to the measurement which is independent of the genetic value so that there is no 'interaction' between genotype and environment. This 'environmental deviation' is supposed to be normally distributed with zero mean and constant variance, and is not correlated among relatives. Then, measuring from the mean, we can write

$$\begin{aligned} x &= \text{observed value} \\ &= y \text{ (genetic value)} + \text{environmental effect} \\ &= z \text{ (representative value)} + \text{dominance deviation} + \text{environmental effect.} \end{aligned}$$

These three terms, the first two of which are sums over the various loci, are mutually uncorrelated. Thus with a large number of loci, the joint distribution of (x, y, z) is trivariate normal, with z (representative value), $y-z$ (dominance deviation), and $x-y$ (environmental effect) all statistically independent. It therefore follows that

$$\begin{aligned} \operatorname{cov}(x, y) &= \operatorname{var}(y) = V, \\ \operatorname{cov}(x, z) &= \operatorname{cov}(y, z) = \operatorname{var}(z), \\ \operatorname{var}(x) &= \operatorname{var}(y) + \operatorname{var}(\eta), \end{aligned}$$

where η is the environmental effect.

Thus for the regression coefficients we find

$$b_{x.y} = b_{x.z} = b_{y.z} = 1.$$

Then an increase δz in the representative value will on the average increase both the genetic component y, and the observed measurement z, by δz. This is also evident from the above decomposition.

Thus we have

$$b_{y.x} = c_1 \text{ (say)} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} = \frac{\operatorname{var}(y)}{\operatorname{var}(x)},$$

and, using (XXVIb),

$$b_{z.y} = c_2 \text{ (say)} = \frac{\operatorname{var}(z)}{\operatorname{var}(y)} = \frac{\tau^2}{\sigma^2 - A\epsilon^2}.$$

Now let x, y, z be the values for a father, and X, Y, Z, the corresponding values for his son. The regressions of the values X, Y, Z, on x, y, z will arise in two ways. In the first place, the partial regression of Z on z, keeping the mother fixed, will be $\frac{1}{2}$ (from the table in section 5).

The dominance deviations $(y-z)$, $(Y-Z)$, and the environmental effects $(x-y)$, $(X-Y)$, are uncorrelated with each other and with z, Z. Thus it is easy to find the regressions of any of X, Y, Z, on any of x, y, z.

However, there is a second indirect component of regression arising from the fact that the son's Z is correlated with the mother's representative value, which is in turn correlated with the father's because of assortative mating. Fisher now finds this extra component of regression under three different hypotheses about the nature of assortative mating, namely that the underlying association is between (1) the observed characters x; (2) the genetic components y; (3) the representative values z.

Notice that he now uses μ for the observed correlation between the x values, whereas in the previous discussion it was the correlation between the y's.

" 17. The effects both of dominance and of environment may be taken into account in calculating the coefficient of correlation: if we call x the actual height of the individual, y what his height would have been under some standard environment, and z what his height would have been if in addition, without altering the extent to which different factors are associated, each phase is given its representative value of Article 5. Then, since we are using the term environment formally for arbitrary external causes independent of heredity, the mean x of a group so chosen that $y = t$ for each member will be simply t, but the mean y of a group so chosen that $x = t$ for each member will be $c_1 t$, where c_1 is a constant equal to the ratio of the variance with environment absolutely uniform to that when difference of environment also makes its contribution. Similarly for the group $z = t$, the mean value of y is t, but for the group $y = t$ the mean z is $c_2 t$, where

$$c_2 = \frac{\tau^2}{\sigma^2 - Ae^2}. \qquad \text{(XXVII)}$$

" Now, we may find the parental and grandparental correlations from the fact that the mean z of any sibship is the mean z of its parents: but we shall obtain very different results in these as in other cases, according to the interpretation which we put upon the observed correlation between parents. For, in the first place, this correlation may be simply the result of conscious selection. If the correlation for height stood alone this would be the most natural interpretation. But it is found that there is an *independent* association of the length of the forearm:* if it is due to selection it must be quite unconscious, and, as Professor Pearson points out, the facts may be explained if to some extent fertility is dependent upon genetic similarity. Thus there are two possible interpretations of marital correlations. One regards the association of the apparent characteristics as primary: there must, then, be a less intense association of the genotype y, and still less of z. The other regards the association as primarily in y or z, and as appearing somewhat masked by environmental effects in the observed correlation. In the first place, let us suppose the observed correlation in x to be primary."

In the discussion below, assuming this first interpretation of marital correlation, if one parent has the value $x = t$, the children will have the value

$$c_1 c_2 \frac{1+\mu}{2} t$$

and not

$$c_1 c_2 \frac{1+\mu}{2}$$

as misprinted in the paper. The remainder of the formulae follow.

" Then if μ is the correlation for x, $c_1 \mu$ will be that for y, and this must be written for μ in the applications of the preceding paragraphs. Hence

$$A = c_1 c_2 \mu,$$

and μ, $c_1 \mu$ and A are the marital correlations for x, y, and z.

" Since the mean z of a sibship is equal to the mean z of its parents, we may calculate the parental and grandparental correlations thus: for group chosen so that $x = t$: mean y, $\bar{y} = c_1 t$; mean z, $\bar{z} = c_1 c_2 t$; $\bar{x}$ of mate is μt; $\bar{z}$ of mate is $c_1 c_2 \mu t$. Therefore $\bar{z}$ of children is

$$c_1 c_2 \frac{1+\mu}{2}.$$

" Hence, since there is no association except of z between parents and child, the parental correlation coefficient is

$$c_1 c_2 \frac{1+\mu}{2}.$$

Now, since we know the mean z of the children to be

$$c_1 c_2 \frac{1+\mu}{2} t,$$

the mean z of their mates is

$$c_1 c_2 \frac{1+\mu}{2} At,$$

* Pearson and Lee, ' On the Laws of Inheritance in Man.' *Biometrika*, **2**, 374.

and the grandparental correlation coefficient will be

$$c_1 c_2 \frac{1+\mu}{2} \frac{1+A}{2}.$$

Similarly, that for the $(n+1)$th parent will be

$$c_1 c_2 \frac{1+\mu}{2} \left(\frac{1+A}{2}\right)^n,$$

giving the Law of Ancestral Heredity as a necessary consequence of the factorial mode of inheritance.

" 18. If we suppose, on the other hand, that the association is essentially in y, the coefficient of correlation between y of husband and y of wife must be μ/c_1 in order to yield an apparent correlation μ. Also

$$c_2 = \frac{\tau^2}{\sigma_2 - A\epsilon^2},$$

and

$$A = \frac{\mu}{c_1} c_2.$$

μ is the observed correlation of x's. If the structural correlation occurs in the y's, it must therefore have value μc_1^{-1} so that

$$\mu = c_1(\mu c_1^{-1})$$

and the argument proceeds as before.

" The parental correlation found as before is now

$$\frac{c_1 c_2 + A c_1}{2},$$

and the higher ancestors are given by the general form

$$\frac{c_1 c_2 + A c_1}{2} \left(\frac{1+A}{2}\right)^n,$$

although A is now differently related to c_1, c_2 and μ.

" In the third case, where the essential connection is between z of husband and z of wife—and this is a possible case if the association is wholly due to selective fertility or to the selection of other features affected by the same factors—the equation between the correlations for y and z is changed, for now the marital correlation for y is equal to Ac_2 when we retain the definition

$$c_2 = \frac{\tau^2}{\sigma^2 - A\epsilon^2}.$$

" Hence also

$$\mu = A c_1 c_2,$$

and the correlation coefficients in the ancestral line take the general form

$$c_1 c_2 \left(\frac{1+A}{2}\right)^{n+1}.$$

" 19. On the first of these theories a knowledge of the marital and the parental correlations should be sufficient to determine $c_1 c_2$, and thence to deduce the constant ratio of the ancestral coefficients.

Thus for three human measurements:

	Stature	Span	Forearm
μ	0·2804	0·1989	0·1977
p	0·5066	0·4541	0·4180
$c_1 c_2$	0·7913	0·7575	0·6980
A	0·2219	0·1507	0·1377
$\frac{1}{2}(1+A)$	0·6109	0·5753	0·5689

These figures are deduced from those given by Pearson and Lee (*loc. cit.*), neglecting sex distinctions, which are there found to be insignificant, and taking the weighted means."

In the table above, μ is the observed correlation between mates as taken from Pearson and Lee, and p is the observed parental–offspring correlation. We then find $c_1 c_2$, A, and $\frac{1}{2}(1+A)$ from the formulae

$$p = c_1 c_2 \frac{1+\mu}{2}, \quad A = c_1 c_2 \mu.$$

"These values for $\frac{1}{2}(1+A)$ agree very satisfactorily with the two ratios of the ancestral correlations which have been obtained, 0·6167 for eye colour in man, and 0·6602 for coat colour in horses. It is evident that if we also knew the ratio of the ancestral correlations for these features, we could make a direct determination of A and ascertain to what extent it is the cause and to what extent an effect of the observed marital correlation.

"20. The correlations for sibs, double cousins, and more distant relations of the same type, in which all the ancestors of a certain degree are common, may be found by considering the variance of the group of collaterals descended from such ancestors. The variance of a sibship, for example, depends, apart from environment, only upon the number of factors in which the parents are heterozygous, and since the proportion of heterozygotes is only diminished by a quantity of the second order, the mean variance of the sibships must be taken for our purposes to have the value appropriate to random mating,

$$\tfrac{1}{2}\tau^2 + \tfrac{3}{4}\epsilon^2 = \tfrac{1}{4}V[2c_2(1-A)+3(1-c_2)]$$

plus the quantity $(V/c_1)-V$ due to environment. But the variance of the population is V/c_1; and the ratio of the two variances must be $1-f$, where f is the fraternal correlation. Hence

$$f = \tfrac{1}{4}c_1(1+c_2+2c_2A)."$$

Still assuming the first model of correlation basically between the x's, we have to find the 'variance of a sibship'. We imagine the number of individuals in a sibship indefinitely increased, and then the x's of the resulting individuals will have a distribution with mean m_s, say, and variance v_s. Both of these will depend on the genetic character of the parents. The observed value, x, of a random sib from a random sibship may be decomposed into two parts as

$$x = m_s + (x - m_s),$$

where x and m_s are both random variables. Since in any one sibship we have

$$E(x-m_s) = 0,$$

by definition, we also must have $E\{m_s(x-m_s)\} = 0$

within each sibship, and therefore in the whole population. Thus m_s and $(x-m_s)$ are uncorrelated. From this it follows that

$$\begin{aligned}\text{var}\,(x) &= \text{var}\,(m_s) + \text{var}\,(x-m_s),\\ &= v_x, \text{ say.}\end{aligned}$$

Here var $(x-m_s)$ means the mean value of $(x-m_s)^2$ taken over all sibs in all sibships. It is therefore the mean value of v_s taken over all sibships and can be written as $\bar{v}_s$. Then

$$\text{var}\,(m_s) = v_x - \bar{v}_s.$$

If x, X, are the measurements of a random pair of sibs from a random sibship,

$$\begin{aligned}\text{cov}\,(x, X) &= \text{cov}\,(m_s + \{x-m_s\}, m_s + \{X-m_s\})\\ &= \text{var}\,(m_s)\\ &= v_x - \bar{v}_s.\end{aligned}$$

Thus the sib–sib correlation is

$$f = \frac{\text{cov}\,(x,X)}{\sqrt{\{\text{var}\,(x)\,\text{var}\,(X)\}}} = \frac{v_x - \bar{v}_s}{\sqrt{\{v_x v_x\}}} = 1 - \frac{\bar{v}_s}{v_x}.$$

This can be written $$\bar{v}_s = (1-f)\,v_x.$$

The variance, v_s, within any sibship depends only on segregation within that sibship and therefore only on those genes for which the parents are heterozygous, since if the parents are homozygous the effect is to make a constant addition to all sibs alike. But the frequencies of heterozygotes at any locus are affected by assortative mating only by a small quantity so that the variance within sibships will be changed by a proportionally small quantity. Thus $\bar{v}_s$ can be taken, nearly enough, to have its value for random mating, although var (m_s) will have to be changed.

If there are no environmental effects, and no assortative mating, the correlation between the sibs is $$\frac{\tau^2+\frac{1}{2}\epsilon^2}{2\sigma^2}.$$

Thus the covariance between sibs is $\frac{1}{2}\tau^2+\frac{1}{4}\epsilon^2$,

which will be unaffected by any environmental effects which are such that they are uncorrelated in the sibs. We also have

$$V = \text{var}\,(y) = \sigma^2+\frac{A}{1-A}\tau^2$$

$$= \tau^2+\epsilon^2+\frac{A}{1-A}\tau^2,$$

$$c_2 = \frac{\tau^2}{\sigma^2-A\epsilon^2} = \frac{\tau^2}{\tau^2+(1-A)\epsilon^2}.$$

Solving these equations for τ^2 and ϵ^2 we get

$$\tau^2 = Vc_2(1-A), \quad \epsilon^2 = V(1-c_2)$$

From these we have

$$\text{cov}\,(x, X) = \tfrac{1}{2}\tau^2+\tfrac{1}{4}\epsilon^2$$
$$= \tfrac{1}{4}V\{2c_2(1-A)+3(1-c_2)\}.$$

We also have $$c_2 = \text{var}\,(y)/\text{var}\,(x) = Vv_x^{-1},$$

and substituting in the formula for f we get Fisher's result.

For double cousins we argue as follows. At any one locus each member of a double cousinship may be regarded as having one gene chosen at random from the four carried by his father's parents, and one chosen at random from the four carried by his mother's parents. The variances of the cousins within the cousinship will depend only on the dissimilarities within each of these two sets of four genes, and therefore by the same argument as before, will be almost independent of assortative mating.

Let x and X be the observed values of the two double cousins, and f the correlation between them. The variance of the population, and therefore of x or X is Vc_1^{-1}, and the variance due to environmental effects is $Vc_1^{-1}-V_1$. Then the variance of $x-X$ must be

$$E(x-X)^2 = 2Vc_1^{-1}(1-f)$$

on the one hand, and $$E(x-X)^2 = 2V(c_1^{-1}-1)+2(\sigma^2-\tfrac{1}{4}\tau^2-\tfrac{1}{16}\epsilon^2)$$

on the other, because the correlation between the genetic components for double cousins is known to be $$\frac{1}{4\sigma^2}(\tau^2+\tfrac{1}{4}\epsilon^2).$$

Thus the second term above is the genetic component of variance.

Putting $\sigma^2 = \tau^2 + \epsilon^2$, and substituting for τ^2 and ϵ^2, we get

$$Vc_1^{-1}(1-f) = V(c_1^{-1}-1) + V\{\tfrac{3}{4}c_2(1-A) + \tfrac{15}{16}(1-c_2)\},$$

so that

$$1-f = 1-c_1 + \tfrac{3}{4}c_1c_2(1-A) + \tfrac{15}{16}c_1(1-c_2),$$

and

$$f = c_1\{\tfrac{1}{16} + \tfrac{3}{16}c_2 + \tfrac{3}{4}Ac_2\}.$$

" In the same way, the variance for a group of double cousins is unaffected by selective mating, and we find the correlation coefficient for double cousins to be

$$\tfrac{1}{16}c_1(1 + 3c_2 + 12c_2A),$$

showing how the effect of selective mating increases for the more distant kin.

" On the first hypothesis, then, we must write,

$$\mu = \frac{A}{c_1c_2}, \quad p = c_1c_1\frac{1+\mu}{2},$$

and

$$f = \tfrac{1}{4}c_1\{1 + c_2(1+2A)\}.$$

" 21. We shall use this formula for the fraternal correlation to estimate the relative importance of dominance and environment in the data derived from the figures given by Pearson and Lee.

" Assuming as the observed correlations

	Stature	Span	Cubit
μ	0·2804	0·1989	0·1977
p	0·5066	0·4541	0·4180
f	0·5433	0·5351	0·4619

we obtain as before

c_1c_2	0·7913	0·7575	0·6980
A	0·2219	0·1507	0·1377

and calculating c_1 from the formula $c_1 = 4f - c_1c_2(1+2A)$,

we obtain the three values 1·031 1·155 0·957

with a standard error of 0·072, and a mean of 1·048."

Presumably by 'standard error' Fisher means 'standard deviation of the observed values'. However, this is not clear; the standard deviation based on two degrees of freedom would be 0·100, not 0·072 and the standard errors in the next table also do not agree. It is not clear what precisely is in Fisher's mind here. He does all his calculations to three or four decimal places. But he does not give any indication of the accuracy of the correlations on which his calculations are based, other than the 'standard errors' quoted from time to time. These do not seem to be standard errors in the sense of the term most used nowadays, namely, the standard deviation of the estimate to be expected in repeated sampling. The text suggests that the three values of c_1 for respectively stature, span and cubit were looked upon as if they were three estimates of some 'ideal' or 'true' value of c_1, differing from this only by random fluctuations.

" This relatively large standard error, due principally to our comparative ignorance of the fraternal correlations (errors in μ have scarcely any effect, and those in p are relatively unimportant), prevents us from making on a basis of these results a close estimate of the contributions to the total variance of the factors under consideration.

" Remembering that c_1 is intrinsically less than unity, the second value is inexplicably high, whilst the first and third are consistent with any value sufficiently near to unity. The mean of these results is materially greater than unity, and therefore gives no support to the supposition that there is any cause of variance in these growth features other than genetic differences. If this is so, we should put $c_1 = 1$, and compare the observed values of f with those calculated from the formula

$$4f = 1 + c_2(1 + 2A).$$

" With their standard errors we obtain

	Stature	Span	Cubit	Standard error
Observed	0·5433	0·5351	0·4619	0·016
Calculated	0·5356	0·4964	0·4726	0·008
Difference	−0·0077	−0·0387	+0·0107	0·018

" The exceptional difference in the fraternal correlations for span might, perhaps, be due to the effects of epistacy, or it may be that the terms which we have neglected, which depend upon the finiteness of the number of factors, have some influence. It is more likely, as we shall see, that the assumption of direct sexual selection is not justified for this feature. Accepting the above results for stature, we may ascribe the following percentages of the total variance to their respective causes:

	%	%
Ancestry		54
Variance of sibship:		
$\frac{1}{2}\tau^2$	31	
$\frac{3}{4}\epsilon^2$	15	
Other causes		
		46
		100

Again it may be divided:

Genotypes (σ^2):		
Essential genotypes (τ^2)	62	
Dominance deviations (ϵ^2)	21	
		83
Association of factors by homogamy		17
Other causes		—
		100

" These determinations are subject, as we have seen, to considerable errors of random sampling, but our figures are sufficient to show that, on this hypothesis, it is very unlikely that so much as 5 per cent of the total variance is due to causes not heritable, especially as every irregularity of inheritance would, in the above analysis, appear as such a cause.

" It is important to see that the large effect ascribed to dominance can really be produced by ordinary Mendelian factors. The dominance ratio ϵ^2/σ^2, which may be determined from the correlations, has its numerator and denominator composed of elements, δ^2 and α^2, belonging to the individual factors. We may thereby ascertain certain limitations to which our factors must be subject if they are successfully to interpret the existing results. The values of the dominance ratio in these three cases are found to be:

	Stature	Span	Cubit	Standard error
Dominance ratio	0·253	0·274	0·336	0·045

" 22. The correlations for uncles and cousins, still assuming that the association of factors is due to a direct selection of the feature x, may be obtained by the methods of Article 14, using the two series already obtained: that for ancestors

$$c_1 c_2 \frac{1+\mu}{2}\left(\frac{1+A}{2}\right)^n,$$

and that for collaterals, like sibs and double cousins, which have all their ancestors of a certain degree in common,

$$\tfrac{1}{4}c_1[1 + c_2(1 + 2A)],$$

$$\tfrac{1}{16}c_1[1 + 3c_2(1 + 4A)],$$

and so on.

" Thus if a group be chosen so that $x = t$,

$$\bar{y} \text{ of group is } c_1 t,$$

$$\bar{z} \text{ of group is } c_1 c_2 t,$$

$$z \text{ of sibs is } c_1 c_2 \frac{1+A}{2} t,$$

also

$$\bar{y} \text{ of sibs is } \tfrac{1}{4} c_1 [1 + c_2(1+2A)]\, t,$$

$$\bar{y} \text{ of sibs mates is } \tfrac{1}{4} c_1 [1 + c_2(1+2A)]\, c_1 \mu t,$$

$$\bar{z} \text{ of sibs mates is } \tfrac{1}{4} c_1 [1 + c_2(1+2A)]\, A t.$$

Hence

$$\bar{z} \text{ of nephews is } \tfrac{1}{8} c_1 [2c_2(1+A) + \{1 + c_2(1+2A)\} A]\, t,$$

giving the correlation

$$c_1 c_2 \left(\frac{1+A}{2}\right)^2 + \tfrac{1}{8} c_1 A (1 - c_2).$$

" Again for cousins, if a group be chosen so that $x = t$, we have

$$\bar{y} \text{ of uncles is } \left[c_1 c_2 \left(\frac{1+A}{2}\right)^2 + \tfrac{1}{8} c_1 A (1 - c_2) \right] t,$$

$$\bar{z} \text{ of uncles is } c_1 c_2 \left(\frac{1+A}{2}\right)^2,$$

and

$$\bar{z} \text{ of uncles mates is } \left[c_1 c_2 \left(\frac{1+A}{2}\right)^2 + \tfrac{1}{8} c_1 A (1 - c_2) \right] A t,$$

hence

$$\bar{z} \text{ of cousins is } \left[c_1 c_2 \left(\frac{1+A}{2}\right)^3 + \tfrac{1}{16} c_1 A^2 (1 - c_2) \right] t,$$

giving the correlation

$$c_1 c_2 \left(\frac{1+A}{2}\right)^3 + \tfrac{1}{16} c_1 A^2 (1 - c_2).$$

" The formulae show that these two correlations should differ little from those for grandparent and great-grandparent, using the values already found, and putting $c_1 = 1$ we have

	Stature	Span	Cubit
Grandparent	0·3095	0·2612	0·2378
Great-grandparent	0·1891	0·1503	0·1353
Uncle	0·3011	0·2553	0·2311
Cousin	0·1809	0·1445	0·1288

" 23. On the third supposition, that the marital correlation is due primarily to an association in the essential genotype z, we obtain results in some respects more intelligible and in accordance with our existing knowledge.

" From the fundamental equations

$$\mu = c_1 c_2 A, \quad p = \tfrac{1}{2}(c_1 c_2 + \mu),$$

we may deduce

$$c_1 c_2 = 2p - \mu, \quad A = \mu/(2p - \mu),$$

whence the following table is calculated:

	Stature	Span	Cubit	Standard error
μ	0·2804	0·1989	0·1977	0·0304
p	0·5066	0·4541	0·4180	0·0115
f	0·5433	0·5351	0·4619	0·0160
$c_1 c_2$	0·7328	0·7093	0·6383	0·038
A	0·3826	0·2804	0·3097	0·028
$\frac{1}{2}(1+A)$	0·6913	0·6402	0·6549	0·014

and making use of the fraternal correlations to separate c_1 and c_2, by the equations

$$f = \tfrac{1}{4} c_1 [1 + c_2(1+2A)],$$

or

$$c_1 = 4f - 2p - \mu,$$

we obtain

c_1	0·8796	1·0333	0·8139	0·078
c_2	0·8331	0·6864	0·7842	0·077
e^2/σ^2	0·2450	0·3883	0·2850	0·105

"The standard error for the dominance ratio is now very high, since the latter is proportional to the difference $f-p$. If we assume a known value for c_1, and calculate the dominance ratio from p and μ only, the standard error falls nearly to its value in Article 18.

"The three values for the ratio of the ancestral correlations 0·691, 0·640, 0·655 are now higher than that obtained from observations of eye colour, and are more similar to the value 0·660 obtained for the coat colour of horses. Without knowing the marital correlations in these cases, it is not possible to press the comparison further. It would seem unlikely that the conscious choice of a mate is less influenced by eye colour than by growth features, even by stature. But it is not at all unlikely that eye colour is but slightly correlated with other features, while the growth features we know to be highly correlated, so that a relatively slight selection in a number of the latter might produce a closer correlation in each of them than a relatively intense selection of eye colour.

"The value of c_1 for span is still greater than unity, 1·033, but no longer unreasonably so, since the standard error is about 0·078. If we were considering span alone the evidence would be strongly in favour of our third hypothesis. A remarkable confirmation of this is that Pearson and Lee (*loc. cit.* p. 375), considering organic and marital correlations alone, show that the observed correlations could be accounted for by the following direct selection coefficients:

Stature	Span	Cubit
0·2374	0·0053	0·1043

Naturally these cannot be taken as final, since there are a large number of other features, which may be connected with these and at the same time may be subject to sexual selection. The correlations of cross assortative mating are in fact smaller than they would be if direct selection to this extent were actually taking place. The influence of other features prevents us from determining what proportion of the observed association is due to direct selection, but if inheritance in these growth features is capable of representation on a Mendelian scheme—and our results have gone far to show that this is likely—it would be possible to distinguish the two parts by comparing the parental and fraternal correlations with those for grandparents and other kindred.

"On our present supposition that the association is primarily in z and for the case of span this seems likely, the correlations for uncle and cousin will be the same as those for grandparent and great-grandparent, being given by the formulae

$$c_1 c_2 \left(\frac{1+A}{2}\right)^2 \quad \text{and} \quad c_1 c_2 \left(\frac{1+A}{2}\right)^3,$$

leading to the numbers

	Stature	Span	Cubit
Grandparent	0·3502	0·2907	0·2737
Great-grandparent	0·2421	0·1861	0·1793

Fisher now considers the hypothesis that the observed correlation μ between the phenotypes x of the parents arises as the summation of two effects. The first is a direct correlation s, which is the result of direct sexual selection. Fisher calls this the 'coefficient of selection'. The second part, $\mu - s$, is a reflection of a correlation between their z-values, arising differently. Each of these parts can be treated as regression coefficient. He thus supposes that the effect on a child is the sum of the effects which arise by these two causes.

Now the direct correlation or regression s between the phenotypes x of the parents produces a correlation $c_1 c_2 s$ between their z-values, as shown in Section 22, and hence a regression $c_1 c_2 s$ of the z-value of the father on that of the mother. The further correlation $\mu - s$ between the parents' x-values is a reflection of a correlation $(\mu - s)/c_1 c_2$ between their z-values, as shown in Section 17, and hence a regression $(\mu - s)/c_1 c_2$. If we suppose that these can be legitimately added together, the total regression of one z-value on the other is

$$A = c_1 c_2 s + (\mu - s)/c_1 c_2$$

and this is equal to their correlation.

Similarly the direct correlation s produces a regression $\frac{1}{2}c_1c_2(1+s)$ of child on parent, and the correlation $(\mu-s)/c_1c_2$ between the z-values of the parents produces a further regression $(\mu-s)/2$. Adding these, we find for the total regression of child on parent, which is the same as the correlation between them

$$p = \tfrac{1}{2}c_1c_2(1+s)+\tfrac{1}{2}(\mu-s).$$

The argument by which Fisher deduces the value

$$\begin{aligned} f &= \tfrac{1}{4}c_1(1+c_2+2c_2A) \\ &= \tfrac{1}{4}c_1+\tfrac{1}{4}c_1c_2(1+2A) \end{aligned}$$

for the correlation between sibs still holds. From it we find

$$c_1 = c_1c_2(1+2A)-4f.$$

" 24. Neither these nor the similar table for the first hypothesis accord ill with the value obtained for uncle and nephew, 0·265, from measurements of eye colour. It may, however, be thought that neither of them give high enough value for cousins. Certainly they do not approach some of the values found by Miss Elderton in her memoir on the resemblance of first cousins (*Eugenics Laboratory Memoirs*, IV). Series are there found to give correlations over 0·5, and the mean correlation for the measured features is 0·336. From special considerations this is reduced to 0·270, but if the similarity of first cousins is due to inheritance, it must certainly be less than that between uncle and nephew. No theory of inheritance could make the correlation for cousins larger than or even so large as that for the nearer relationship.

" It will be of interest finally to interpret our results on the assumption that the figures quoted (Article 20) represent actual coefficients of selection. Manifestly it would be better to obtain the value of A experimentally from the ratio of the ancestral correlations, using the collateral correlations to determine what are the marital correlations for y. For the present we must neglect the possibility of an independent selection in y: and although we know that the figures are not final, we shall write s, the coefficient of selection, equal to 0·2374, 0·0053, and 0·1043 in our three cases.

" Further, let

$$A = c_1c_2s+\frac{\mu-s}{c_1c_2},$$

so that

$$2p = c_1c_2(1+s)+\mu-s,$$

whence we deduce

	Stature	Span	Cubit
c_1c_2	0·7841	0·7108	0·6725
A	0·2410	0·2761	0·2090
$\frac{1}{2}(1+A)$	0·6205	0·6381	0·6045

the values of A being now in much closer agreement for the three features. Further, from the fraternal correlation we have

c_1	1·0112	1·0370	0·8940

with a mean at 0·9821.

" Again, for the dominance ratio

0·2763 0·3880 0·2940 0·3194 (mean),

leaving a trifle under 2 per cent for causes not heritable, but requiring high values about 0·32 for the dominance ratio.

" 25. *The Interpretation of the Statistical Effects of Dominance.* The results which we have obtained, although subject to large probable errors and to theoretical reservations which render an exact estimate of these errors impossible, suggest that the ratio ϵ^2/σ^2, the statistical measure of the extent of dominance, has values of about 0·25 to 0·38. In his initial memoir on this subject Karl Pearson has shown that, under the restricted conditions there considered, this ratio should be exactly $\frac{1}{3}$. Subsequently Udny Yule (Conference on Genetics) pointed out that the parental correlation could be raised from the low values reached in that memoir to values more in accordance with the available figures by the partial or total abandonment of the assumption of dominance. To this view Professor Pearson subsequently gave his approval: but it does not seem to have been observed that if lower values are required—and our analysis tends to show that they are not—the statistical effects are governed not only by the physical ratio d/a, but by the proportions in which the three Mendelian phases are

present. This effect is an important one, and very considerably modifies the conclusions which we should draw from any observed value of the dominance ratio.

" The fraction δ^2/α^2, of which the numerator and denominator are the contributions of a single factor to ϵ^2 and σ^2, is equal, as we have seen (Article 5, equations V–VII) to

$$\frac{2pqd^2}{(p+q)^2a^2-2(p^2-q^2)ad+(p^2+q^2)d^2},$$

and depends wholly upon the two ratios d/a and p/q. We may therefore represent the variations of this function by drawing the curves for which it has a series of constant values upon a plane, each point on which is specified by a pair of particular values for these two ratios. The accompanying diagram (fig. 1) shows such a series of curves, using d/a and $\log(p/q)$ as co-ordinates. The logarithm is chosen as a variable, because equal intensity of selection will affect this quantity to an equal extent, whatever may be its value; it also possesses the great advantage of showing reciprocal values of p/q in symmetrical positions."

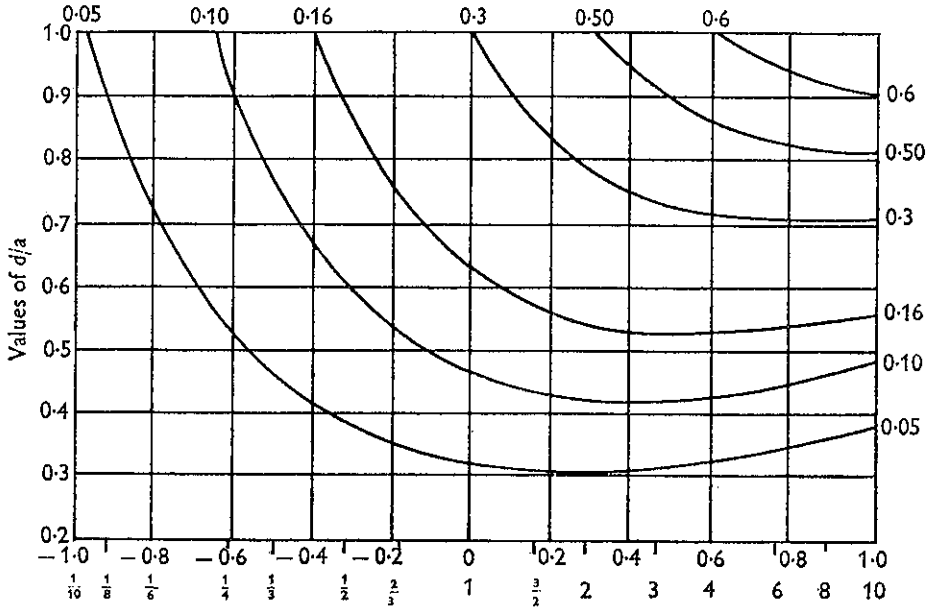

Fig. 1. Values of $\log_{10}(p/q)$ (upper figures) and of p/q (lower figures).

The dominance ratio given above is obtained by simple substitution of $P = p^2$, $Q = pq$, $R = q^2$, $p+q = 1$, into (V) and (VII).

In the paragraph below, the figure 3 is misprinted for 0·3.

" It will be seen that 3 is not by any means the highest value possible: when $d = a$, and when p/q is very great, any value up to unity may appear; but high values are confined to this restricted region. When d/a is less than 0·3 the ratio is never greater than 0·05, and we cannot get values as high as 0·15 unless d/a be as great as 0·5. On the other hand, all values down to zero are consistent with complete dominance, provided that the values of p/q are sufficiently small.

" We know practically nothing about the frequency distribution of these two ratios. The conditions under which Mendelian factors arise, disappear, or become modified are unknown. It has been suggested that they invariably arise as recessive mutations in a dominant population. In that case p/q would initially be very high, and could only be lowered if by further mutation, and later by selection, the recessive phase became more frequent. These factors would, however, have little individual weight if better balanced factors were present, until p/q had been lowered to about 10. In face of these theories it cannot be taken for granted that the distribution of these ratios is a simple one. It is natural, though possibly not permissible, to think of their distributions as independent. We may profitably consider further the case in which the distribution is symmetrical, in which the factor of known a and d is equally likely to be more frequent in the dominant as in the recessive phase.

" For this case we combine the numerators and denominators of the two fractions

$$\frac{2pqd^2}{(p+q)^2a^2-2(p^2-q^2)ad+(p^2+q^2)d^2} \quad \text{and} \quad \frac{2pqd^2}{(p+q)^2a^2+2(p^2-q^2)ad+(p^2+q^2)d^2},$$

and obtain the joint contribution $$\frac{2pqd,}{(p+q)^2a^2+(p^2+q^2)d^2}$$

the curves for which are shown in fig. 2, representing the combined effect of two similar factors, having their phases in inverse proportions. It will be seen that complete dominance does not preclude the possibility of low value for the dominance ratio: the latter might fall below 0·02 if the greater part of the variance were contributed by factors having the ratio between p and q as high as 100 to 1. This ratio is exceedingly high; for such a factor only one individual in 10,000 would be a recessive. We may compare the frequency of deaf mutism with which about one child in 4000 of normal parents is said to be afflicted. It would be surprising if more equal proportions were not more common, and if this were so, they would have by far the greater weight.

" The fact that the same intensity of selection affects the logarithm of p/q equally, whatever its value may be, suggests that this function may be distributed approximately according to the law of errors. This is a natural extension of the assumption of symmetry, and is subject to the same reservations. For instance, a factor in which the dominant phase is the commonest would seem less likely to suffer severe selection than one in which the recessive phase outnumbers the other. But if symmetry be granted, our choice of a variable justifies the consideration of a normal distribution.

" Writing ξ for $\log_e p/q$ and σ for the standard deviation of ξ, we have

$$p = e^{\frac{1}{2}\xi}/2\cosh\tfrac{1}{2}\xi, \quad q = e^{-\frac{1}{2}\xi}/2\cosh\tfrac{1}{2}\xi \quad \text{and} \quad 2pq = \tfrac{1}{2}\operatorname{sech}^2\tfrac{1}{2}\xi.$$

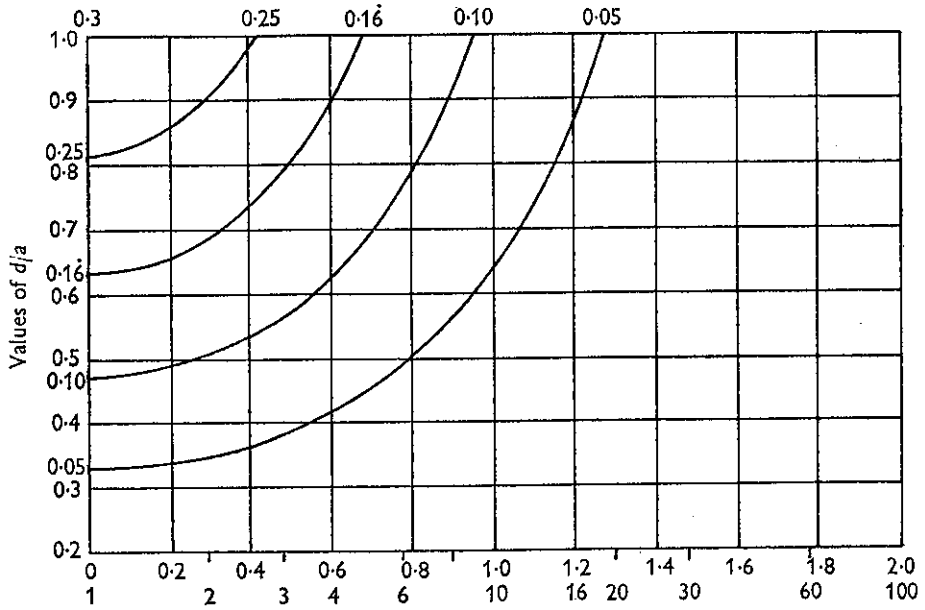

Fig. 2. Values of $\log_{10}(p/q)$ (upper figures) and of p/q (lower figures).

" Hence we have to evaluate

$$E = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} \tfrac{1}{2}\operatorname{sech}^2\tfrac{1}{2}\xi \,.\, e^{-\xi^2/2\sigma^2}d\xi = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \tfrac{1}{2}\operatorname{sech}^2\tfrac{1}{2}\sigma\xi. \quad e^{-\frac{1}{2}\xi^2}d\xi, \qquad \text{(XXVIII)}$$

and the dominance ratio derived from the whole group is

$$\frac{Ed^2}{a^2+(1-E)d^2}.$$

" E is a function of σ only, which decreases steadily from its value $\frac{1}{2}$ when $\sigma = 0$, approaching when σ is large to the function $2/(\sigma\sqrt{2\pi})$. The function $(16+16\sigma^2+\frac{1}{4}\pi^2\sigma^4)^{-\frac{1}{4}}$ osculates it at the origin, and appears on trial to represent it effectively to three significant figures. This function has been used for calculating the form of the accompanying curves. Fig. 3 shows the course of the function E. Fig. 4 gives the curves comparable to those of figs. 1 and 2, showing the value of the dominance ratio for different values d/a and σ. If the assumptions upon which this diagram is based are justified, we are now advanced some way towards the interpretation of an observed dominance ratio. A ratio of 0·25 gives us a lower limit of about 0·8 for d/a, and no upper limit. If the possibility of superdominance ($d > a$) is excluded, then the ratio of the phases must be so distributed

that the standard ratio e^{σ} is not greater than about 3 : 1. A greater value of the standard ratio would make the effect of dominance too small; a smaller value could be counteracted by a slight reduction of d/a. We have therefore no reason to infer from our dominance ratios that dominance is incomplete. We may speak of it as having at least four-fifths of its full value, but we can set no upper limit to it.

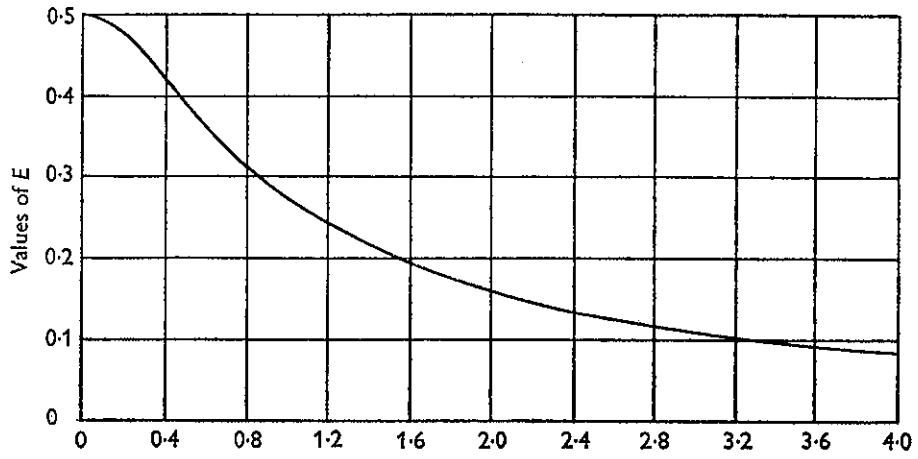

Fig. 3. Values of $\sigma \log_{10} e = \cdot 4343\sigma$.

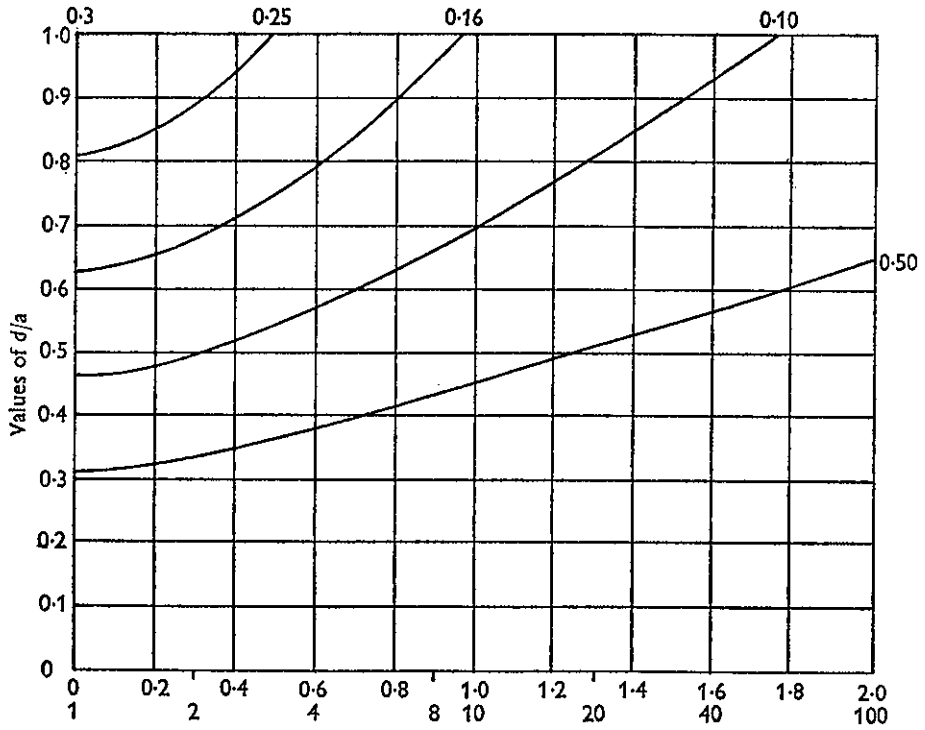

Fig. 4. Values of $\log_{10}$ of standard ratio (upper figures) and of standard ratio (lower figures).

" 26. Throughout this work it has been necessary not to introduce any avoidable complications, and for this reason the possibilities of Epistacy have only been touched upon, and small quantities of the second order have been steadily ignored. In spite of this, it is believed that the statistical properties of any feature determined by a large number of Mendelian factors have been successfully elucidated. Due allowance has been made for the factors differing in the magnitude of their effects, and in their degree of dominance, for the possibility of Multiple Allelomorphism and of one important type of Coupling. The effect of the dominance in the individual factors has been seen to express itself in a single Dominance Ratio. Further the effect of marital correlation has been fully examined, and the relation between this association and the coefficient of marital correlation has been made clear.

" By means of the paternal correlation it is possible to ascertain the dominance ratio and so distinguish dominance from all non-genetic causes, such as environment, which might tend to lower the correlations: this is due to the similarity in siblings of the effects of dominance which causes the fraternal correlation to exceed the parental. The fact that this excess of the fraternal correlation is very generally observed is itself evidence in favour of the hypothesis of cumulative factors. On this hypothesis it is possible to calculate the numerical influence not only of dominance, but of the total genetic and non-genetic causes of variability. An examination of the best available figures for human measurements shows that there is little or no indication of non-genetic causes. The closest scrutiny is invited on this point, not only on account of the practical importance of the predominant influence of natural inheritance, but because the significance of the fraternal correlation in this connection has not previously been realised.

" Some ambiguity still remains as to the causes of marital correlations; our numerical conclusions are considerably affected according as this is assumed to be of purely somatic or purely genetic origin. It is striking that the indications of the present analysis are in close agreement with the conclusions of Pearson and Lee as to the genetic origin of a part of the marital correlation, drawn from the effect of the correlation of one organ with another in causing the selection of one organ to involve the selection of another. This difficulty will, it is hoped, be resolved when accurate determinations are available of the ratio of the grandparental to the parental correlation. From this ratio the degree of genetic association may be immediately obtained, which will make our analysis of the Variance as precise as the probable errors will allow.

" In general, the hypothesis of cumulative Mendelian factors seems to fit the facts very accurately. The only marked discrepancy from existing published work lies in the correlation for first cousins. Snow, owing apparently to an error, would make this as high as the avuncular correlation; in our opinion it should differ by little from that of the great-grandparent. The values found by Miss Elderton are certainly extremely high, but until we have a record of complete cousinships measured accurately and without selection, it will not be possible to obtain satisfactory numerical evidence on this question. As with cousins, so we may hope that more extensive measurements will gradually lead to values for the other relationship correlations with smaller standard errors. Especially would more accurate determinations of the fraternal correlation make our conclusions more exact.

" Finally, it is a pleasure to acknowledge my indebtedness to Major Leonard Darwin, at whose suggestion this inquiry was first undertaken, and to whose kindness and advice it owes its completion."