XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance.

R. A. Fisher

Transactions of the Royal Society of Edinburgh / Volume 52 / Issue 02 / January 1919, pp 399 - 433
DOI: 10.1017/S0080456800012163, Published online: 06 July 2012

XV.—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 June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

CONTENTS.

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 $\sqrt{\sigma_1^2+\sigma_2^2}$. 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 ·5, we have accounted by reference to the height of the father for one quarter of the variance of the sons. 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 ·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.

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

* 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 $\frac{V}{\sigma^2}=1-r$. Taking the values ·5066 and ·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.

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 $\frac{1}{3}$ and the fraternal $\frac{5}{12}$.*

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 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.

* 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 formulæ 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. Edin.*, Jan. 1910).

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

$$P(a-m)^2 + 2Q(d-m)^2 + R(a+m)^2 = a^2 \quad . \quad . \quad . \quad . \quad . \quad (I)$$

a^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 a^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

$$\sigma^2 = \Sigma a^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (II)$$

To justify our statement that a^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

$$M_3 = \Sigma\mu_3$$
$$M_4 - 3\sigma^4 = \Sigma(\mu_4 - 3a^4).$$

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

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

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

2. If there are a great number of different factors, so that σ is large compared to every separate a, 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 heterozygates $x-d+m$, and the recessives $x+a+m$, and since the variance of the original population was σ^2-a^2, the frequencies of these three types are in the ratio

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

or, when σ is great compared to a, so that $\frac{a^2}{\sigma^2}$ may be neglected,

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

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

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

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

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

and therefore the deviation of the mean of the offspring is

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

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

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

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 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

$$\mathrm{P}(c+b-a)^2+2\mathrm{Q}(c-d)^2+\mathrm{R}(c-b+a)^2$$

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

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

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

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

we obtain

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

and since

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

we have

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

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

$$\mathrm{Q}^2=\mathrm{PR},$$

for then

$$\delta^2=4\mathrm{Q}^2d^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{V})$$

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

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

and

$$a^2=2a^2\mathrm{Q}-4\mathrm{Q}(\mathrm{P}-\mathrm{R})ad+2\mathrm{Q}(\mathrm{P}+\mathrm{R})d^2 \quad . \quad . \quad . \quad . \quad (\mathrm{VII})$$

In general

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

and if

$$\sigma^2 = \Sigma a^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{VIII})$$

$$\tau^2 = \Sigma \beta^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{IX})$$

and

$$\epsilon^2 = \Sigma \delta^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{X})$$

then

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

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

$$\frac{x}{\sigma^2} \quad \frac{\tau^2}{2},$$

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

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

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 $\frac{\tau}{\sigma}$. 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 $\frac{\tau^2}{\sigma^2}$.

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

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

so that

$$i\mathrm{P}+2j\mathrm{Q}+k\mathrm{R}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(XII)}$$

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

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

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

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

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	$4p^3q$	$2p^2q^2$	$4p^2q^2$	$4pq^3$	$q4$
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.

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.

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

$p^2(p+\frac{1}{4}q)^2$	$\frac{3}{2}p^2q(p+\frac{1}{4}q)$	$\frac{9}{16}p^2q^2$
$\frac{3}{2}p^2q(p+\frac{1}{4}q)$	$\frac{1}{2}pq(p^2+\frac{13}{2}pq+q^2)$	$\frac{3}{2}pq^2(\frac{1}{4}p+q)$
$\frac{9}{16}p^2q^2$	$\frac{3}{2}pq^2(\frac{1}{4}p+q)$	$q^2(\frac{1}{4}p+q)^2$

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.

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

$\frac{1}{4}p^2(4p+q)$	$\frac{1}{4}p^2q^2(7p+q)$	$\frac{3}{4}p^2q^2$
$\frac{1}{4}p^2q(7p+q)$	$\frac{1}{4}pq(p^2+14pq+q^2)$	$\frac{1}{4}pq^2(p+7q)$
$\frac{3}{4}p^2q^2$	$\frac{1}{4}pq^2(p+7q)$	$\frac{1}{4}q^3(p+4q)$

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

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.

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

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.

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

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

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

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

which also takes the form

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

The parental table

$p^3/4q$	$-\frac{1}{4}pq$
$-\frac{1}{4}pq$	$q^3/4p$

yields

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

and the fraternal table

$p^2/4q$	—
—	$q^2/4p$

leads us to the simple expression

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

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.

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

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

and the frequency in the range dy is

$$\frac{1}{\sqrt{2\pi \mathrm{V}}} e^{-\frac{y^2}{2\mathrm{V}}} dy = \mathrm{N};$$

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

$$\frac{1}{2\pi \mathrm{V}\sqrt{1-\mu^2}} e^{-\frac{1}{1-\mu^2}\frac{x^2-2\mu xy+y^2}{2\mathrm{V}}} dx\, dy,$$

which is equal to

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

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

$$\mathrm{MN} e^{\frac{\mu ab}{\mathrm{V}}}.$$

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 × D and R × R obviously no change is made, and the same is true of the crosses D × H and R × H, for each of these yields the pure form and the heterozygote in equal numbers. On the other hand, in the cross D × R we have a dominant and a recessive replaced in the next generation by two heterozygotes, while in the cross H × 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,

$$4Q^2 e^{\frac{\mu J^2}{V}} = 4PRe^{\frac{\mu IK}{V}}.$$

Since $\frac{J^2}{V}$ and $\frac{IK}{V}$ are small quantities, we shall neglect their squares, and obtain the equation

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

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

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

It is evident that

$$PI + 2QJ + RK = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (XV)$$

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

$$DH' \text{ is } 2PQ'(1 + f_{12}),$$

and that of

$$DR' \text{ is } PR'(1 + f_{13}),$$

and so on.

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

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

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

$$\left.\begin{aligned} \mathrm{I} &= i + \Sigma(\mathrm{P}'i'f_{11} + 2\mathrm{Q}'j'f_{12} + \mathrm{R}'k'f_{13}) \\ \mathrm{J} &= j + \Sigma(\mathrm{P}'i'f_{21} + 2\mathrm{Q}'j'f_{22} + \mathrm{R}'k'f_{23}) \\ \mathrm{K} &= k + \Sigma(\mathrm{P}'i'f_{31} + 2\mathrm{Q}'j'f_{32} + \mathrm{R}'k'f_{33}) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \text{(XVII)}$$

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

$$\mathrm{V} = \Sigma(\mathrm{P}i^2 + 2\mathrm{Q}j^2 + \mathrm{R}k^2) + 2\Sigma\{\mathrm{PP}'(1 + f_{11})ii' + 8 \text{ other terms}\},$$

which reduces to

$$\Sigma(\mathrm{P}i^2 + 2\mathrm{Q}j^2 + \mathrm{R}k^2) + 2\Sigma\{\mathrm{PP}'ii'f_{11} + 8 \text{ other terms}\},$$

so that

$$\mathrm{V} = \Sigma(\mathrm{P}i\mathrm{I} + 2\mathrm{Q}j\mathrm{J} + \mathrm{R}k\mathrm{K}) \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(XVIII)}$$

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 $\mathrm{DD}' \times \mathrm{DD}'$ have the frequency

$$\mathrm{PP}' \,.\, \mathrm{PP}' \,.\, (1 + f_{11})(1 + f_{11}) e^{\frac{\mu(\mathrm{I}+\mathrm{I}')^2}{\mathrm{V}}},$$

which for our purpose is equal to

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

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

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

Now since

$$(\mathrm{P} + \mathrm{Q})^2(\mathrm{P}' + \mathrm{Q}')^2 - \mathrm{PP}'(\mathrm{P} + 2\mathrm{Q} + \mathrm{R})(\mathrm{P}' + 2\mathrm{Q}' + \mathrm{R}') = (\mathrm{Q}^2 - \mathrm{PR})\mathrm{P}' + (\mathrm{Q}'^2 - \mathrm{P}'\mathrm{R}')\mathrm{P} + (\mathrm{Q}^2 - \mathrm{PR})(\mathrm{Q}'^2 - \mathrm{P}'\mathrm{R}')$$

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

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

Also collecting the terms in I and J, we find

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

which yields on eliminating J,

$$\frac{\mu}{4V}[p'(IP-KR)+p(I'P'-K'R')]^2,$$

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

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

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 versâ*, 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

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

so that the set of four equations

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

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

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

where

$$A(1-A)=\frac{\mu}{V}\Sigma(i'P'-k'R')[p'(i'-j')+q'(j'-k')]$$

$$=\frac{\mu}{V}\Sigma\beta^2, \text{ since } \beta^2=\frac{(iP-kR)^2}{2Q}$$

or

$$A(1-A)=\mu\frac{\tau^2}{V} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(XXII)}$$

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

$$\left.\begin{aligned} p^2p'^2f_{11} &= \frac{\mu}{(1-A)^2}\frac{(i\mathrm{P}-k\mathrm{R})(i'\mathrm{P}'-k'\mathrm{R}')}{\mathrm{V}}pp' \\ p^2q'^2f_{12} &= -\frac{\mu}{(1-A)^2}\frac{(i\mathrm{P}-k\mathrm{R})(i'\mathrm{P}'-k'\mathrm{R}')}{\mathrm{V}}pq' \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad \text{(XXIII)}$$

Hence

$$\begin{aligned} \mathrm{I} &= i + \frac{\mu}{(1-A)^2}\frac{i\mathrm{P}-k\mathrm{R}}{p\mathrm{V}}\Sigma[p'(i'-j')+q'(j'-k')](i'\mathrm{P}'-k'\mathrm{R}') \\ &= i + \frac{\mu}{(1-A)^2}\frac{i\mathrm{P}-k\mathrm{R}}{p}\cdot\frac{\tau^2}{\mathrm{V}}, \end{aligned}$$

and so

$$\left.\begin{aligned} \mathrm{I} &= i + \frac{A}{1-A}\frac{i\mathrm{P}-k\mathrm{R}}{p} \\ \text{Similarly}\quad \mathrm{K} &= k - \frac{A}{1-A}\frac{i\mathrm{P}-k\mathrm{R}}{q} \\ \text{and}\quad \mathrm{J} &= j - \frac{A}{1-A}\frac{p-q}{2pq}(i\mathrm{P}-k\mathrm{R}) \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \text{(XXIV)}$$

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

$$\mathrm{V} = \Sigma(\mathrm{P}i\mathrm{I} + 2\mathrm{Q}j\mathrm{J} + \mathrm{R}k\mathrm{K})$$

reduces to a similar form. For evidently

$$\mathrm{V} = \Sigma a^2 + \frac{A}{1-A}\cdot\Sigma(i\mathrm{P}-k\mathrm{R})[p(i-j)+q(j-k)].$$

Hence

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

Therefore the equation for A finally takes the form

$$\mu\tau^2 = \mathrm{V}A(1-A) = A(1-A)\sigma^2 + A^2\tau^2,$$

and may be otherwise written

$$A^2\epsilon^2 - A\sigma^2 + \mu\tau^2 = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(XXVI)}$$

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

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

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

$$\frac{A}{\mu} = \frac{\tau^2}{\sigma^2 - A\epsilon^2} = \frac{\tau^2 + \frac{A}{1-A}\tau^2}{\sigma^2 + \frac{A}{1-A}\tau^2} \quad . \quad . \quad . \quad . \quad . \quad \text{(XXVI, } b\text{)}$$

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, . . . , and the heterozygous phases to $2pq$, $2pr$, $2qr$, . . .

Hence, our measurements being from the mean,

$$p^2i_1+q^2i_2+r^2i_3+\ldots+2pqj_{12}+2prj_{13}+\ldots=0 \quad . \quad . \quad . \quad . \quad \text{(XII*)}$$

As before, we define a^2 by the equation

$$p^2i_1^2+q^2i_2^2+r^2i_3^2+\ldots+2pqj_{12}^2+2prj_{13}^2+\ldots=a^2 \quad . \quad . \quad . \quad . \quad \text{(I*)}$$

and choosing l, m, n, . . . , 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

$$\beta^2=\mathrm{S}(4l^2p^2)+\mathrm{S}(2pq\overline{l+m}^2),$$
$$=\mathrm{S}(2p(1+p)l^2)+\mathrm{S}(4pqlm),$$

and since

$$pl+qm+rn+\ldots=0,$$
$$\beta^2=\mathrm{S}(2pl^2),$$

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

$$2p^3i_1^2+4p^2qi_1j_{12}+2pq(p+q)j_{12}^2+4pqrj_{12}j_{13}.$$

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.

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

$$\tfrac{1}{4}p^2(1+p)^2{i_1}^2+\tfrac{1}{2}p^2q^2i_1i_2+p^2q(1+p)i_1j_{12}+p^2qri_1j_{13}$$
$$\tfrac{1}{2}pq(1+p+q+2pq){j_{12}}^2+pqr(1+2p)j_{12}j_{13}+2pqrsj_{12}j_{34},$$

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

$$\tfrac{1}{4}p^2(1+2p){i_1}^2+p^2qi_1j_{12}+\tfrac{1}{2}pq(1+p+q){j_{12}}^2+pqrj_{12}j_{13},$$

or, simply,

$$\tfrac{1}{4}(\alpha^2+\beta^2).$$

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, . . . be the mean deviations of the homozygous phases, and J_{12}, J_{13}, . . . 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

$$p\mathrm{I}_1+q\mathrm{J}_{12}+r\mathrm{J}_{13}+\ldots=\mathrm{L},$$
$$p\mathrm{J}_{12}+q\mathrm{I}_2+r\mathrm{J}_{23}+\ldots=\mathrm{M},$$

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

$$2pq(1+f_{12})=2pqe^{\frac{\mu}{\mathrm{V}}\cdot\mathrm{LM}},$$

that is,

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

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

$$\begin{aligned}&pp'f''_{11\,.\,11}+pq'f''_{11\,.\,12}+pr'f''_{11\,.\,13}+\ldots\\&+qp'f''_{12\,.\,11}+qq'f''_{12\,.\,12}+qr'f''_{12\,.\,13}+\ldots\end{aligned}$$

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

$$\begin{aligned}2f''_{12\,.\,12}&=\mathrm{F}_{11}+\mathrm{F}_{22}+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{L}')(\mathrm{M}+\mathrm{M}')\\&\quad+\mathrm{F}_{12}+\mathrm{F}_{21}+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{M}')(\mathrm{M}+\mathrm{L}'),\end{aligned}$$

but

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

therefore

$$2f_{12\,.\,12}=\mathrm{F}_{11}+\mathrm{F}_{22}+\mathrm{F}_{12}+\mathrm{F}_{21}+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{M})(\mathrm{L}'+\mathrm{M}') \quad . \quad . \quad . \quad \textbf{(XIX*)}$$

By analogy with Article 12, the solution

$$f_{12\,.\,12}=\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{M})(\mathrm{L}'+\mathrm{M}')$$

suggests itself, and on trial it leads to

$$\mathrm{F}_{11}=\frac{\mu}{\mathrm{V}}\mathrm{L}\mathrm{L}',$$

and is thereby verified.

Hence we may evaluate L, L', . . . , for

$$\begin{aligned}\mathrm{L}&=p\mathrm{I}_1+q\mathrm{J}_{12}+r\mathrm{J}_{13}+\ldots\\&=l+\Sigma\{p'^2i'(pf_{11\,.\,11}+qf_{12\,.\,11}+\ldots)+2p'q'j'_{12}(pf_{11\,.\,12}+qf_{12\,.\,12}+\ldots)+\ldots\end{aligned}$$

but

$$pf_{11\,.\,11}+qf_{12\,.\,11}+\ldots=\frac{\mu}{\mathrm{V}}\mathrm{L}(\mathrm{L}'+\mathrm{M}'),$$

therefore

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

Let

$$L = l + AL,$$

then

$$L = \frac{l}{1-A},$$

and

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

therefore

$$A(1-A) = \frac{\mu}{V}\Sigma(2p'l'^2 + 2q'm'^2 + \ldots)$$

$$= \frac{\mu}{V}\Sigma\beta'^2,$$

therefore

$$A(1-A) = \frac{\mu}{V}\tau^2 \quad \ldots\ldots\ldots\ldots \quad \text{(XXII*)}$$

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

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

and that the variance reduces, as before, to

$$\sigma^2 + \frac{A}{1-A}\tau^2 \quad \ldots\ldots\ldots \quad \text{(XXV*)}$$

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 $(1\,.\,1) \times (2\,.\,2)$ or $(1\,.\,2) \times (2\,.\,1)$; 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).

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

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

Further, if these gametic types occur with frequency,

$$\mathfrak{p}=pp'\left(1+\frac{\mu}{\mathrm{V}}\mathrm{LL}'\right)\quad \mathfrak{q}=pq'\left(1+\frac{\mu}{\mathrm{V}}\mathrm{LM}'\right)$$

$$\mathfrak{r}=qp'\left(1+\frac{\mu}{\mathrm{V}}\mathrm{ML}'\right)\quad \mathfrak{s}=qq'\left(1+\frac{\mu}{\mathrm{V}}\mathrm{MM}'\right),$$

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

$$p^2p'^2(1+f''_{11\,.\,11})=p^2p'^2\left\{1+\frac{\mu}{\mathrm{V}}(\mathrm{L}^2+\mathrm{L}'^2+4\mathrm{LL}')\right\}$$

$$\mathfrak{p}^2\left\{1+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{L}')^2\right\}=\mathfrak{p}^2\left(1+\frac{\mu}{\mathrm{V}}\mathfrak{L}^2\right),$$

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

$$2p^2p'q'(1+f''_{11\,.\,12})=2p^2p'q'\left\{1+\frac{\mu}{\mathrm{V}}\mathrm{L}^2+\mathrm{L}'\mathrm{M}'+2\mathrm{L}(\mathrm{L}'+\mathrm{M}'))\right\}$$

$$=2\mathfrak{pq}\left\{1+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{L}')(\mathrm{L}+\mathrm{M}')\right\}=2\mathfrak{pq}\left(1+\frac{\mu}{\mathrm{V}}\mathfrak{L}\mathfrak{M}\right).$$

Finally, taking half the double heterozygotes,

$$2pqp'q'(1+f''_{12\,.\,12})=2pqp'q'\left\{1+\frac{\mu}{\mathrm{V}}(\mathrm{LM}+\mathrm{L}'\mathrm{M}'+(\mathrm{L}+\mathrm{M})(\mathrm{L}'+\mathrm{M}'))\right\}$$

$$2\mathfrak{ps}\left\{1+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{L}')(\mathrm{M}+\mathrm{M}')\right\}=2\mathfrak{ps}\left(1+\frac{\mu}{\mathrm{V}}\mathfrak{L}\mathfrak{O}\right),$$

or, equally,

$$2\mathfrak{qr}\left\{1+\frac{\mu}{\mathrm{V}}(\mathrm{L}+\mathrm{M}')(\mathrm{M}+\mathrm{L}')\right\}=2\mathfrak{qr}\left(1+\frac{\mu}{\mathrm{V}}\mathfrak{M}\mathfrak{N}\right).$$

From this it 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.

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_1t, 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_2t, where

$$c_2=\frac{\tau^2}{\sigma^2-A\epsilon^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(XXVII)}$$

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.

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

$$A=c_1c_2\mu,$$

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, $\bar{y}=c_1t$; mean z, $\bar{z}=c_1c_2t$; $\bar{x}$ of mate is μt; $\bar{z}$ of mate is $c_1c_2\mu t$. Therefore $\bar{z}$ of children is

$$c_1c_2\frac{1+\mu}{2}.$$

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

$$c_1c_2\frac{1+\mu}{2}.$$

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

$$c_1c_2\frac{1+\mu}{2}t,$$

* PEARSON and LEE, "On the Laws of Inheritance in Man," *Biometrika*, ii, 374.

the mean z of their mates is

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

and the grandparental correlation coefficient will be

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

Similarly, that for the $(n+1)^{\text{th}}$ parent will be

$$c_1 c_2 \frac{1+\mu}{2} \left(\frac{1+\mathrm{A}}{2}\right)^n,$$

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 $\frac{\mu}{c_1}$ in order to yield an apparent correlation μ. Also

$$c_2 = \frac{\tau^2}{\sigma^2 - \mathrm{A}\epsilon^2},$$

and

$$\mathrm{A} = \frac{\mu}{c_1} c_2.$$

The parental correlation found as before is now

$$\frac{c_1 c_2 + \mathrm{A}c_1}{2},$$

and the higher ancestors are given by the general form

$$\frac{c_1 c_2 + \mathrm{A}c_1}{2} \left(\frac{1+\mathrm{A}}{2}\right)^n,$$

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 $\mathrm{A}c_2$ when we retain the definition

$$c_2 = \frac{\tau^2}{\sigma^2 - \mathrm{A}\epsilon^2}.$$

Hence also

$$\mu = \mathrm{A}c_1 c_2,$$

and the correlation coefficients in the ancestral line take the general form

$$c_1 c_2 \left(\frac{1+\mathrm{A}}{2}\right)^{n+1}.$$

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	·2804	·1989	·1977
p	·5066	·4541	·4180
c_1c_2	·7913	·7575	·6980
A	·2219	·1507	·1377
$\frac{1}{2}(1+\mathrm{A})$ . . .	·6109	·5753	·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.

These values for $\frac{1}{2}(1+\mathrm{A})$ agree very satisfactorily with the two ratios of the ancestral correlations which have been obtained, ·6167 for eye colour in man, and ·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,

$$\tfrac{1}{2}\tau^2+\tfrac{3}{4}\epsilon^2=\frac{\mathrm{V}}{4}[2c_2(1-\mathrm{A})+3(1-c_2)]$$

plus the quantity $\dfrac{\mathrm{V}}{c_1}-\mathrm{V}$ 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

$$f=\frac{c_1}{4}(1+c_2+2c_2\mathrm{A}).$$

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

$$\frac{c_1}{16}(1+3c_2+12c_2\mathrm{A}),$$

showing how the effect of selective mating increases for the more distant kin.

On the first hypothesis, then, we must write,

$$\mu=\frac{\mathrm{A}}{c_1c_2},$$

$$p=c_1c_2\frac{1+\mu}{2},$$

and

$$f=\frac{c_1}{4}(1+c_2(1+2\mathrm{A})).$$

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	·2804	·1989	·1977
p	·5066	·4541	·4180
f	·5433	·5351	·4619

we obtain as before

c_1c_2	·7913	·7575	·6980
A	·2219	·1507	·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 ·957

with a standard error of ·072, and a mean of 1·048.

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	·5433	·5351	·4619	·016
Calculated	·5356	·4964	·4726	·008
Difference	−·0077	−·0387	+·0107	·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 per cent.
Variance of sibship :		
$\frac{1}{2}\tau^2$	31 per cent.	
$\frac{3}{4}\epsilon^2$	15 „	
Other causes	...	
		46 „
		100 per cent.

Again it may be divided :

Genotypes (σ^2) :		
Essential genotypes (τ^2) . .	62 per cent.	
Dominance deviations (ϵ^2) . .	21 „	
		83 per cent.
Association of factors by homogamy		17 „
Other causes		...
		100 per cent.

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, $\frac{\epsilon^2}{\sigma^2}$, which may be determined from the correlations, has its numerator and denominator composed of elements, δ^2 and a^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 . . .	·253	·274	·336	·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 $x = t$,

$\bar{y}$ of group is $c_1 t$,
$\bar{z}$ of group is $c_1 c_2 t$,
$\bar{z}$ of sibs is $c_1 c_2 \frac{1+A}{2} t$,

also

$\bar{y}$ of sibs is $\frac{1}{4} c_1 [1 + c_2(1+2A)] t$,
$\bar{y}$ of sibs mates is $\frac{1}{4} c_1 [1 + c_2(1+2A)] c_1 \mu t$,
$\bar{z}$ of sibs mates is $\frac{1}{4} c_1 [1 + c_2(1+2A)] A t$.

Hence

$\bar{z}$ of nephews is $\frac{1}{8} c_1 [2c_2(1+A) + \{1 + c_2(1+2A)\} A] t$,

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 formulæ 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 . . .	·3095	·2612	·2378
Great-grandparent . .	·1891	·1503	·1353
Uncle	·3011	·2553	·2311
Cousin	·1809	·1445	·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

$$\mu = c_1 c_2 A,$$
$$p = \tfrac{1}{2}(c_1 c_2 + \mu),$$

we may deduce

$$c_1 c_2 = 2p - \mu,$$
$$A = \mu/(2p - \mu),$$

whence the following table is calculated :—

	Stature.	Span.	Cubit.	Standard Error.
μ . .	·2804	·1989	·1977	·0304
p . .	·5066	·4541	·4180	·0115
f . .	·5433	·5351	·4619	·0160
$c_1 c_2$ . .	·7328	·7093	·6383	·038
A . .	·3826	·2804	·3097	·028
$\frac{1}{2}(1+A)$.	·6913	·6402	·6549	·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 . .	·8796	1·0333	·8139	·078
c_2 . .	·8331	·6864	·7842	·077
$\frac{\epsilon^2}{\sigma^2}$ . .	·2450	·3883	·2850	·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 ·691, ·640, ·655 are now higher than that obtained from observations of eye colour, and are more similar to the value ·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 ·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.
·2374	·0053	·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 formulæ

$$c_1c_2\left(\frac{1+A}{2}\right)^2 \quad \text{and} \quad c_1c_2\left(\frac{1+A}{2}\right)^3,$$

leading to the numbers

	Stature.	Span.	Cubit.
Grandparent . . .	·3502	·2907	·2737
Great-grandparent . .	·2421	·1861	·1793

24. Neither these nor the similar table for the first hypothesis accord ill with the value obtained for uncle and nephew, ·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 ·5, and the mean correlation for the measured features is ·336. From special considerations this is reduced to ·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 ·2374, ·0053, and ·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	·7841	·7108	·6725
A	·2410	·2761	·2090
$\frac{1}{2}(1+A)$ . . .	·6205	·6381	·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	·8940

with a mean at ·9821.

Again, for the dominance ratio

·2763 ·3880 ·2940 ·3194 (mean),

leaving a trifle under 2 per cent. for causes not heritable, but requiring high values about ·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 $\frac{\epsilon^2}{\sigma^2}$, the statistical measure of the extent of dominance, has values of about ·25 to ·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 $\frac{d}{a}$, 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 $\frac{\delta^2}{a^2}$, 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 $\frac{d}{a}$ and $\frac{p}{q}$. 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, p. 430) shows such a series of curves, using $\frac{d}{a}$ and $\log \frac{p}{q}$ 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 $\frac{p}{q}$ in symmetrical positions.

It will be seen that ·3 is not by any means the highest value possible: when $d = a$, and when $\frac{p}{q}$ is very great, any value up to unity may appear; but high values are confined to this restricted region. When $\frac{d}{a}$ is less than ·3 the ratio is never greater than ·05, and we cannot get values as high as ·15 unless $\frac{d}{a}$ be as great as ·5. On the

other hand, all values down to zero are consistent with complete dominance, provided that the values of $\frac{p}{q}$ 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 $\frac{p}{q}$ 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 $\frac{p}{q}$ 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

$$\frac{2pqd^2}{(p+q)^2a^2-2(p^2-q^2)ad+(p^2+q^2)d^2} \text{ and } \frac{2pqd^2}{(p+q)^2a^2+2(p^2-q^2)ad+(p^2+q^2)d^2},$$

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 ·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 $\frac{p}{q}$ 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.

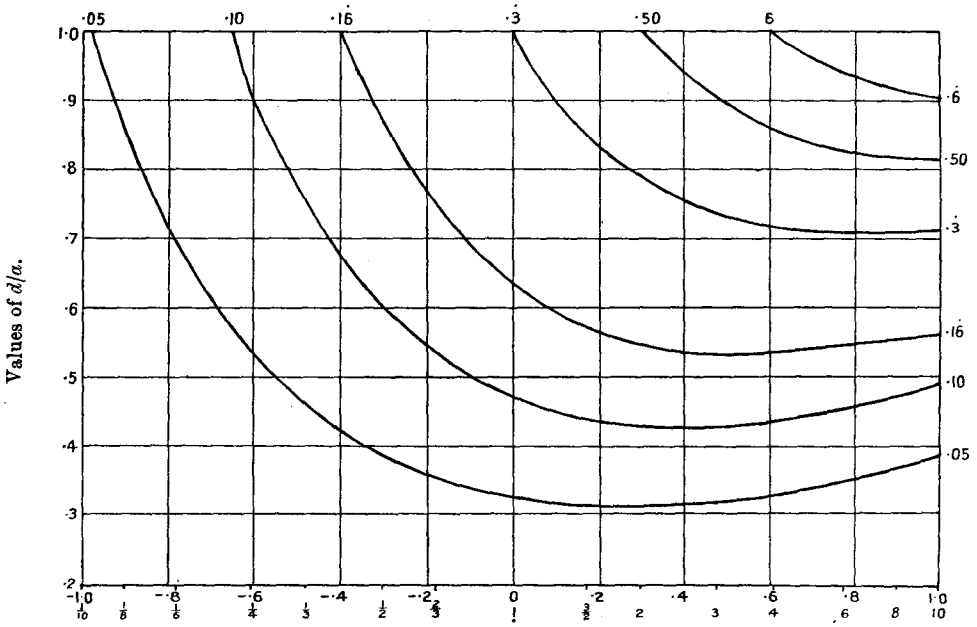

FIG. 1.—Values of $\log_{10}(p/q)$ (upper figures) and of p/q (lower figures).

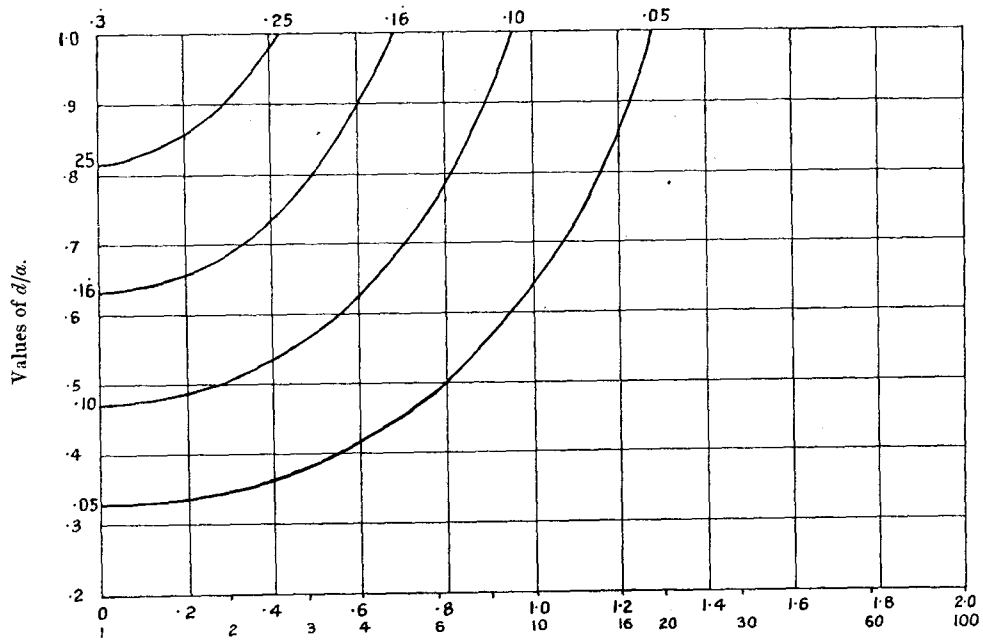

FIG. 2.—Values of $\log_{10}(p/q)$ (upper figures) and of p/q (lower figures).

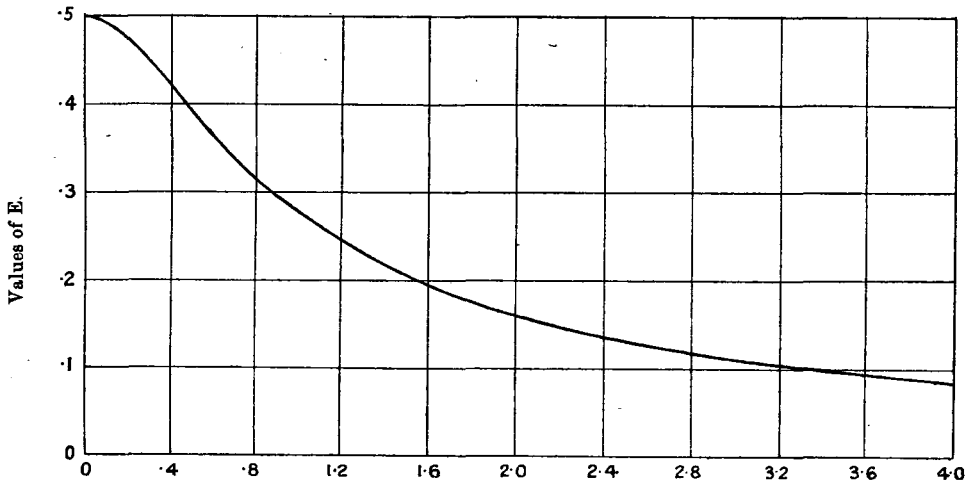

FIG. 3.—Values of $\sigma \log_{10} e = \cdot 4343\sigma$.

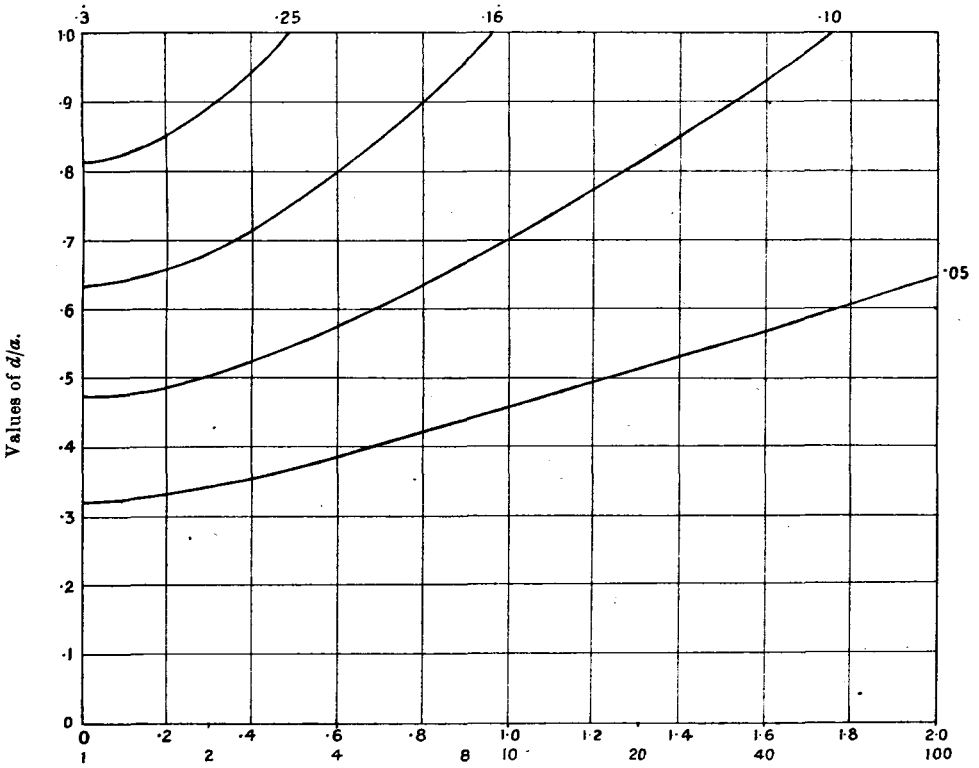

FIG. 4.—Values of $\log_{10}$ of standard ratio (upper figures) and of standard ratio (lower figures).

Writing ξ for $\log_e \frac{p}{q}$ and σ for the standard deviation of ξ, we have

$$p = e^{\frac{1}{2}\xi}/2\cosh\tfrac{1}{2}\xi, \quad q = e^{-\frac{1}{2}\xi}/2\cosh\tfrac{1}{2}\xi, \quad \text{and} \quad 2pq = \tfrac{1}{2}\operatorname{sech}^2\tfrac{1}{2}\xi.$$

Hence we have to evaluate

$$\mathrm{E} = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\tfrac{1}{2}\operatorname{sech}^2\tfrac{1}{2}\xi \,.\, e^{-\xi^2/2\sigma^2}d\xi = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\tfrac{1}{2}\operatorname{sech}^2\frac{\sigma\xi}{2}e^{-\frac{1}{2}\xi^2}d\xi \quad . \quad . \quad \text{(XXVIII)}$$

and the dominance ratio derived from the whole group is

$$\frac{\mathrm{E}d^2}{a^2 + (1-\mathrm{E})d^2}.$$

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 $\frac{2}{\sigma\sqrt{2\pi}}$. The function $\left(16 + 16\sigma^2 + \frac{\pi^2}{4}\sigma^4\right)^{-\frac{1}{4}}$ 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 $\frac{d}{a}$ 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 ·25 gives us a lower limit of about ·8 for $\frac{d}{a}$, 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 $\frac{d}{a}$. 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.

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 fraternal 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.