

*On the Ancestral Gametic Correlations of a Mendelian Population  
mating at Random.*

By KARL PEARSON, F.R.S.

(Received April 2,—Read April 22, 1909.)

(1) The population to be considered in this paper is supposed to be initiated by a group of  $s_1$  individuals with the protogenic constitution (AA),  $s_2$  individuals with the allogenic constitution (aa), and  $s_3$  individuals with the hybrid constitution (Aa), where the mating is given by the simple Mendelian formula:  $(AA) \times (aa) = 4(Aa)$ . I do not assume at this stage any relation between the gametic constitution of an individual and its somatic character. I propose first to consider the correlation between any ancestor and the resulting array of offspring, when we regard only their gametic constitutions. I assume that all mating in the population is random, *i.e.* that every possible mating occurs simply in the proportions of the frequency of individuals of given gametic constitution in the population, and that there is no differential fertility or selective death-rate.

In a paper published in the 'Phil. Trans.,' vol. 203, A, 1904, p. 53 *et seq.*, I have dealt with the correlation between the *somatic* characters of the ancestry and the offspring in a population of a Mendelian character, more general in that I supposed the character to depend upon  $n$  couplets, and not a single Mendelian couplet, less general in that I supposed the population to have arisen from a series of initial hybridisations, and not from a mixture as in the present case of hybrids and members of two pure races in any proportions. In that paper I showed (*a*) that there was correlation between any ancestor and the offspring, (*b*) that the regression for any ancestor and the offspring was linear, and (*c*) that the correlations decreased in geometrical progression. These are the chief characteristics of the Law of Ancestral Heredity. It was clear that, judged by somatic characters only, ancestry was of importance. The result depended on Mendel's first principle of dominance being absolutely true. The values of the correlations were, however, less than those with which biometric work had made us familiar.

(2) In the present paper I start with a more general population, and investigate the correlation of the gametic not the somatic characters.

The general formula for the population before the first mating is

$$s_1(AA) + 2s_3(Aa) + s_2(aa). \quad (i)$$

After the first random mating it is

$$(s_1 + s_3)^2(AA) + 2(s_1 + s_3)(s_2 + s_3)(Aa) + (s_2 + s_3)^2(aa).$$

I write this for brevity

$$p^2(AA) + 2pq(Aa) + q^2(aa), \quad (ii)$$

and this constitution remains permanent in all successive matings. Hence the standard deviations of the gametic constitutions remain the same generation after generation, and the correlation coefficient is in every case equal to the slope of the regression line. I shall determine the slope of this line which will give the correlation and show that the regression is truly linear in each case.

(3) I consider first the effect of individuals of each special type mating with the general population (ii).

- (a) Type (AA): the array of offspring is  $(p+q)[p(AA)+q(Aa)]$ ,  
 (b) Type (aa): „ „  $(p+q)[p(Aa)+q(aa)]$ ,  
 (c) Type (Aa): „ „  $\frac{1}{2}(p+q)[p(AA)+(p+q)(Aa)+q(aa)]$ .

Thus, in seeking what any differentiated group

$$t_1(AA) + t_3(Aa) + t_2(aa)$$

produces when mated with the general population, *i.e.* when mated at random, all we have to do is to replace (AA), (Aa) and (aa) by the above three expressions respectively.

In this manner I obtained the array of offspring due to any parent, any grandparent and any great grandparent. These at once allowed me to reach the general law of distribution, and, assuming this, one multiplication by the general population (ii) demonstrated by induction the validity of the results reached. These are as follows:—

I term  $n$ th parent any individual  $n$  generations back in the direct ancestry: thus a 1st parent is the father or mother; a 2nd parent, a grandparent; a 3rd parent, a great grandparent, and so on.

(i) If the  $n$ th parent be an (AA), then the array of offspring due to random matings is

$$p^2(p+q)^{2(n-1)}\left(\frac{1}{2}\right)^{n-1}\{(2^{n-1}p+q)p(AA)+[(2^n-1)p+q]q(Aa) \\ + (2^{n-1}-1)q^2(aa)\}.$$

(ii) If the  $n$ th parent be an (Aa), then the array of offspring is

$$(pq)(p+q)^{2(n-1)}\left(\frac{1}{2}\right)^{n-1}\{[(2^n-1)p+q]p(AA)+[p^2+2(2^n-1)pq+q^2](Aa) \\ + [p+(2^n-1)q]q(aa)\}.$$

(iii) If the  $n$ th parent be an (aa), then the array of offspring is

$$q^2(p+q)^{2(n-1)}\left(\frac{1}{2}\right)^{n-1}\{(2^{n-1}-1)p^2(AA)+[(2^n-1)q+p]p(Aa) \\ + (2^{n-1}q+p)q(aa)\}.$$

(4) These distributions correspond to the cases of 2, 1 and 0 A elements in the gametic constitution of the  $n$ th parent. And we have at once the following result:—

Number of protogenic elements in $n$ th parent.	Average number of same elements in array of offspring.
2 .....	$\frac{2^{n+1}p + 2q}{2^n(p+q)} = \bar{y}_2,$
1 .....	$\frac{(2^{n+1}-1)p + q}{2^n(p+q)} = \bar{y}_1,$
0 .....	$\frac{(2^{n+1}-2)p}{2^n(p+q)} = \bar{y}_0.$

Accordingly, the average number of protogenic elements in the array of offspring decreases uniformly with the decrease in number of the like elements in the  $n$ th parent, *i.e.*

$$\bar{y}_2 - \bar{y}_1 = (\frac{1}{2})^n = \bar{y}_1 - \bar{y}_0.$$

Thus the regression between the  $n$ th parent and the offspring is linear, and the correlation coefficients form a geometrical series of ratio  $\frac{1}{2}$ , and first term  $\frac{1}{2}$ . Further, the exact constitution of the population, as far as the number of protogenic, allogenic or heterogenic individuals is concerned, is of no influence on the result at all. For all mixtures following the simple Mendelian rule:  $(AA) \times (aa) = 4(Aa)$ , the ancestral correlations for gametic constitution are:

Parental correlation .....	0.500
Grandparental correlation .....	0.250
Great grandparental correlation.....	0.125 and so on.

It will be seen at once that these correlations are of the type  $p, p^2, p^3$ , etc., for which, in my memoir of 1896, I worked out the multiple regression formula, and showed that the ancestors were quite indifferent. "A knowledge of the ancestry beyond the parents in no way alters our judgment as to the size of organ or degree of characteristic probable in the offspring nor its variability."\* This remark and the proof apply equally of course to gametic and to somatic characters if the correlation be of the above form.

(5) Accordingly there remains not the least antinomy between the Mendelian theory and the Law of Ancestral Heredity, if we confine our attention to gametic constitution. The Mendelian ancestry is correlated with the offspring in a series descending in a geometrical progression, and the regression is linear. The values of the correlation coefficients are

\* "Regression, Heredity, and Panmixia," 'Phil. Trans.,' A, vol. 187, 1896, p. 306.

precisely those which it was pointed out in 1896 would lead to a knowledge of the parental constitution\* replacing that of the ancestry.

(6) The striking point, however, of the present investigation is that the values now shown theoretically to exist for the ancestral *gametic* correlations in a simple Mendelian mixture are very close to those determined for *somatic* characters in biometric investigations, whereas the *somatic* correlations for a Mendelian population, if we maintain intact the principle of absolute dominance, appear theoretically to be too low.

Thus the value for parental correlation in man, horse, dog and cattle is about 0.5, and for the grandparental correlation lies between 0.25 and 0.30; but this tendency in the grandparent to some slight excess on the Mendelian gametic value must not be given too much weight.

(7) It seems desirable to consider how far the results in my paper of 1904 for the somatic correlations are modified if we assume for our population

$$p^2(AA) + 2pq(Aa) + q^2(aa),$$

and do not make  $p = q$ .

Assuming the principle of dominance to be absolute, I enquire what is the proportion of offspring possessing the dominant character† (*i.e.* (AA) or (Aa)) supposing the  $n$ th parent to possess it (*i.e.* to be (AA) or (Aa)); and again, what is the proportion possessing the dominant character, supposing the  $n$ th parent does not possess it (*i.e.* to be  $aa$ ).

Percentage of dominant offspring.

$$n\text{th parent dominant in somatic character} \dots 100 \times \frac{2^{n-1}p(p+2q)^2 + q^3}{2^{n-1}(p+q)^2(p+2q)},$$

$$n\text{th parent recessive in somatic character} \dots 100 \times \frac{2^{n-1}p(p+2q)^2 - pq(p+2q)}{2^{n-1}(p+q)^2(p+2q)}.$$

From this it follows that the correlation which is equal to the regression is

$$\frac{1}{2^{n-1}} \frac{q}{p+2q}.$$

If  $p = q$ , this is  $\frac{1}{3} \frac{1}{2^{n-1}}$ , in agreement with the conclusion of my memoir

of 1904. But unless  $q/(p+2q) = \frac{1}{2}$ , *i.e.* the number of pure dominants in the population be vanishingly small (as well, of course, as the number of impure dominants!), this is not a series to which the form  $\rho, \rho^2, \rho^3 \dots$  applies, and when we judge (as we must in most instances in man) by the somatic and not the unknown gametic constitution, the ancestry does matter.

\* As a matter of fact, a knowledge of the gametic constitution of the ancestry in any generation would be equally sufficient with that of the parents.

† It is assumed that A is dominant over a.

The following table illustrates the percentages of dominant characterized offspring when we selected an ancestor of given character :—

Ancestor.	Percentage of dominants in offspring.					
	$p = 2q.$		$p = q.$		$q = 2p.$	
	Dominant.	Recessive.	Dominant.	Recessive.	Dominant.	Recessive.
Parent .....	91·7	66·7	83·3	50·0	57·8	33·3
Grandparent .....	90·3	77·8	79·2	62·5	56·7	44·4
3rd parent .....	89·6	83·3	77·1	68·7	56·1	50·0
4th parent .....	89·2	86·1	76·0	71·9	55·8	52·8
5th parent .....	89·1	87·5	75·5	73·4	55·7	54·2
6th parent .....	89·0	88·2	75·3	74·2	55·6	54·9
$\infty$ th parent.....	88·9	88·9	75·0	75·0	55·6	55·6

It will be clear that the difference of the percentage of dominants in the offspring according as a parent, grandparent or great grandparent was dominant or recessive in somatic character is quite marked ; and only as we approach the higher ancestry, where the correlation is growing very weak, does the percentage difference grow imperceptible.

(8) That ancestry does not matter if we know the gametic constitution of the parents, that it does matter if we only know the somatic character of the parents appears to be the solution of one of the difficulties which some have found between the Mendelian and biometric methods of approaching the subject.

There is, however, I venture to think, another aspect of these results which is worthy of fuller consideration. Namely, the fairly close accordance now shown for the first time to exist between the ancestral gametic correlations in a Mendelian population and the observed ancestral somatic correlations suggests that the accordance between gametic and somatic constitutions is for at least certain characters possibly more intimate than is expressed by an absolute law of dominance. If ( $Aa$ ) were a class, or possibly on a wider determinantal theory a group of several classes, marked by an individual somatic character—not invariably identical with the somatic character of ( $AA$ )—there would be little left of contradiction between biometric and Mendelian results as judged by populations sensibly mating at random. It is the unqualified assertion of the principle of dominance which appears at present as the stumbling block.