SYSTEMS OF MATING. I. THE BIOMETRIC RELATIONS BETWEEN PARENT AND OFFSPRING

SEWALL WRIGHT

Bureau of Animal Industry, United States Department of Agriculture, Washington, D. C.

Received October 29, 1920

TABLE OF CONTENTS

INTRODUCTION

Investigations on the mechanism of heredity have progressed to such a point that there need be little hesitation in accepting the recent statement of EAST and JONES (1919), that "Mendelian heredity has proved to be the heredity of sexual reproduction; the heredity of sexual reproduction is Mendelian."

It has also become clear, however, that most cases of inheritance are far from exhibiting the simplicity which MENDEL was fortunate enough to find in certain variations of the pea. Among our domestic animals, for example, specific Mendelian factors have only been demonstrated in the case of a few color variations and a very small number of morphological differences. The great bulk of the variations both in characters of economic importance and in fancy points can only be interpreted as Mendelian on the assumption that each character is affected by a number of independent unit factors. The same is probably true of most varietal and specific differences in nature. The results of SUMNER's (1918) experiments with geographic races of Peromyscus are a good illustration. Moreover, even where there is simplicity on the genetic side, it may be obscured by variation due to external conditions.

It is thus important to understand the consequences to be expected from the Mendelian mechanism when various systems of mating are followed in a population in which the varying characters are determined by factors which are individually beyond control.

CONDITIONS FOR EQUILIBRIUM

It has been shown by PEARSON (1904) that an F_2 population, composed of 25 percent AA, 50 percent Aa and 25 percent aa, remains true to these proportions under random mating. HARDY (1908) showed that there is equilibrium under random mating if the number of heterozygotes is twice the square root of the product of the two homozygous classes, i.e., if the three classes are in the proportions $x^2 AA : 2xy Aa : y^2 aa$. Here x and y are the relative proportions of gametes A and a in the whole population. Where triple allelomorphs are present in the proportions x, y and z, the proportion of the zygotes, when in equilibrium, can be found from the expansion of $(x + y + z)^2$, and so on for larger numbers of allelomorphs. In cases involving multiple factors, the above formulae hold true of course for each separate factor. As there can be no equilibrium until all of the factors are combined at random (assuming random mating), whether there is linkage or not, the proportions of the various combinations of factors under equilibrium can easily be calculated. With three factors, for example, the proportions of the various zygotic formulae can be found from the expansion of $(x + y)^2 (x' + y')^2 (x'' + y'')^2$.

In the case of one factor, equilibrium is reached in the first generation after random mating commences, regardless of the initial composition of the population (WENTWORTH and REMICK 1916). This is not the case with combinations of two or more factors. Here there is in general a constant approach toward an equilibrium point which theoretically is never reached. The approach is naturally slower in the presence of linkage than with independent factors. JENNINGS (1917) gave formulae for deriving the composition of successive generations in the two-factor case. ROBBINS (1918 a) expressed the results in a simpler form, from which the composition of any required generation can be found immediately. He uses p_n, q_n, s_n, and t_n for the proportions of the gametes AB, Ab, aB and ab, respectively. He assumes a gametic ratio of $r : 1 : 1 : r$ due to linkage.

Letting $\Delta_n = q_n s_n - p_n t_n$ he obtains the formulae:

$$p_n = p_0 + \Delta_0 \left[1 - \left(\frac{r}{1+r}\right)^n\right]$$

$$q_n = q_0 - \Delta_0 \left[1 - \left(\frac{r}{1+r}\right)^n\right]$$

$$s_n = s_0 - \Delta_0 \left[1 - \left(\frac{r}{1+r}\right)^n\right]$$

$$t_n = t_0 + \Delta_0 \left[1 - \left(\frac{r}{1+r}\right)^n\right]$$

The zygotic composition can be derived at once from the gametic composition on the basis of random mating of the gametes as shown by JENNINGS. The more complicated formulae for unequal linkage in the two sexes are also given by ROBBINS.

In the problems to be dealt with in the present series of papers, we will assume that the original population is already in equilibrium under random mating. The question to be investigated in each case is the change in composition of the population brought about by a given system of mating. In some cases a new equilibrium point is approached, in others there is no equilibrium until complete homozygosis is reached.

SYSTEMS OF MATING

The primary classification of systems of mating depends on whether we are dealing with material that is known or unknown genetically. Under the former head come all cases in which it has been possible to isolate specific Mendelian factors. Cases can also be included in which, while the individual factors are not known, it is known that they are homozygous. Thus each of the pure breeds of live stock is doubtless homozygous, or nearly so, in many factors. Pure-breeding and the crossing of pure breeds for one generation are systems of mating which give uniform results within limits.

In dealing with material which is unknown genetically, the two most important methods of control are the mating of animals which for one reason or another are like (or unlike) each other, and the bringing about of a differential productivity among the classes.

The similarity between mated animals may be primarily genetic and only incidentally, if at all, somatic, as in the mating of near relatives or mating within local races. On the other hand there may be assortative mating based on somatic resemblance, in which case there is only incidentally, if at all, genetic similarity. Differential productivity may be due to intentional elimination of certain classes, differential death rate, mating rate or fertility.

The effects of continuous self-fertilization were given by JENNINGS (1912). The results of continued brother-sister mating have been investigated by PEARL (1913, 1914 a, 1914 b), FISH (1914), JENNINGS (1914) and ROBBINS (1917). JENNINGS (1916) has also given formulae for various forms of parent-offspring mating. Simple cases of assortative mating and selection have been discussed by JENNINGS (1916) and by WENTWORTH and REMICK (1916). ROBBINS (1918 *c*) has discussed a form of disassortative mating in which brother-sister matings are avoided.

The usual method of attack has been to write out the different zygotic formulae, or in some cases the gametes, involved in the population, determine the changes in the relative proportions of the different classes during a number of generations under the given system of mating and attempt to deduce from this a general formula. It is evident that this method becomes very cumbersome in dealing with more than two factors. It is also difficult to deal with assortative mating which is not perfect or with the effects of consanguine matings which are more remote than between brother and sister.

ANALYSIS BY PATH COEFFICIENTS

These difficulties, however, can be met in part by a different method of attack. The method to be used here is that of path coefficients. This method which is described in detail in another paper (WRIGHT 1921), gives a means of calculating the degree to which a given effect is determined by each of a number of causes in a complex system of correlated variables. It depends on the combination of knowledge of the degrees of correlation among the variables with such knowledge as may be possessed of the causal relation.

Figure 1 is meant to illustrate a system in which the variations of two quantities X and Y are determined in part by independent causes, such as A and D, and in part by common causes, such as B and C. These common causes may be correlated with each other through more remote causes which are not represented in the figure. We will assume that all of the relations are approximately linear and that the influences of the various causes are combined approximately by addition. The path coefficient, measuring the importance of a given path of influence from cause to effect, is defined as the ratio of the standard deviation of the effect when all causes are constant except the one in question, the variability of which is kept unchanged, to the total standard deviation. A path coefficient differs from a coefficient of correlation in having direction.

The symbol $p_{X \cdot A}$ means the coefficient for the path of influence from A to X. In most cases in the present paper, however, it will be more convenient to represent the path coefficients by single letters.

It can be shown that the squares of the path coefficients measure the degree of determination by each cause. If the causes are independent of each other, the sum of the squared path coefficients is unity. If the causes are correlated, terms representing joint determination must be recognized. The complete determination of X in figure 1 by factor A and the correlated factors B and C can be expressed by the equation:

$$a^2 + b^2 + c^2 + 2\,bc\,r_{BC} = 1 \tag{1}$$

The squared path coefficients and the expressions for joint determination measure the portion of the squared standard deviation of the effect due to the causes singly and jointly, respectively.

The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all of the paths by which the variables are connected. In figure 1, X and Y are connected by four paths.

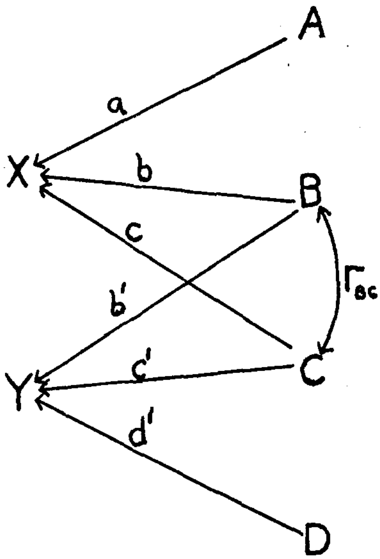

FIGURE 1.—A diagram illustrating the case of two variables (X and Y) determined in part by causes in common (B and C) which are correlated with each other.

$$r_{XY} = bb' + cc' + br_{BC}c' + cr_{BC}b' \tag{2}$$

The mode of analysis of systems of mating, which is to be used here, consists in the main in representing each system by a diagram and applying the two equations given above, one expressing the complete determination of each variable by others, and the other expressing the correlations in terms of path coefficients.

HEREDITY AND ENVIRONMENT

The characteristics of an individual may be looked upon as determined by two classes of factors—those which are internal or hereditary and those which are external or environmental in a broad sense. In figure 2, H

represents the constitution of the fertilized egg and hence all that is received by the individual by inheritance. The environmental factors are separated into two elements, tangible environment (E) and the intangible factors (D) which are not common even to litter mates, and yet appear to be responsible for much variation in early development. The path coefficients $p_{O \cdot H}$, $p_{O \cdot E}$ and $p_{O \cdot D}$ are represented by h, e and d respectively. From equation (1) we have the following equation which is of use in calculating the relative importance of heredity and environment:

$$h^2 + d^2 + e^2 = 1$$

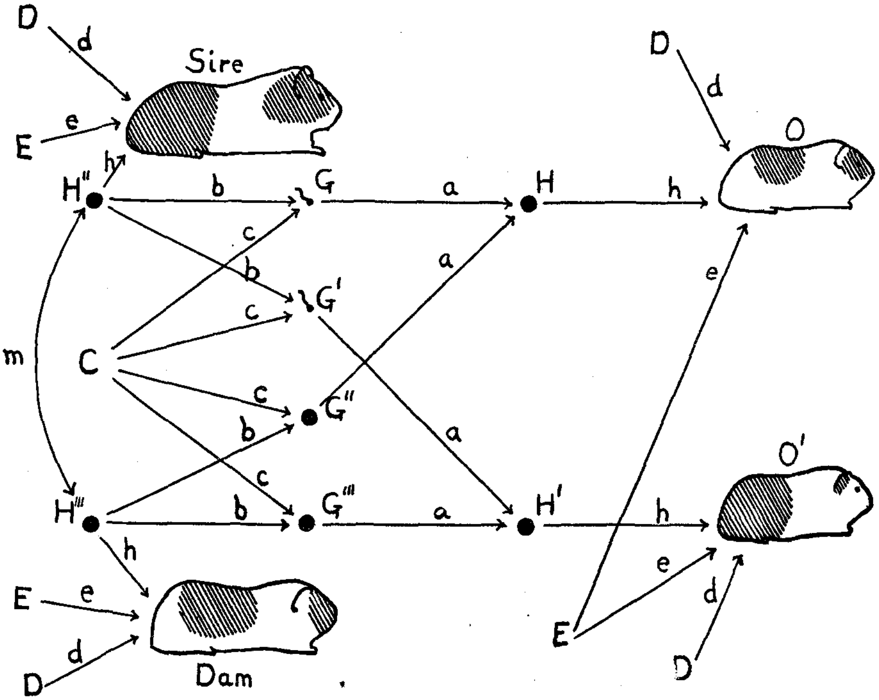

FIGURE 2.—A diagram illustrating the relations between two mated individuals and their progeny. H, H', H'' and H''' are the genetic constitutions of the four individuals. G, G', G'' and G''' are four germ-cells. E and D represent tangible external conditions and chance irregularities as factors in development. C represents chance at segregation as a factor in determining the composition of the germ-cells. Path coefficients are represented by small letters.

THE GRADING OF GAMETIC AND ZYGOTIC FORMULAE

In applying the methods of correlation and of path coefficients to Mendelian inheritance, we must adopt scales of measurement for gametic and zygotic formulae, as well as for the physical characters in question. This would be easy to do if the effects of factors were always combined simply by addition. This, however, is by no means the case. It is probably more typical for the effects to be combined by multiplication, i.e., each factor increases the effect by a certain percentage. This, however, is not a serious difficulty, since it can be met by the use of logarithmic scales if necessary. Cases in which factors produce no effect except in the presence of others are more serious. It seems necessary to ignore them for the present. The most important remaining difficulty is that of dominance. If we assign values to each factor, we get a scale for measuring gametes and zygotes, but the latter represents the effect of the combination in development only if dominance is lacking. For example, if A and B are factors of equal importance, our scale would bring the zygotes $AAbb$, $AaBb$ and $aaBB$ into the same class, while if A and B are dominant, $AaBb$ would show twice as great an effect as the others.

DOMINANCE

This difficulty can be met by using two methods of grading the constitution of the fertilized egg, one which is simply the sum of the values assigned to the various factors and one which measures the effect on development, i.e., allows for dominance. We can find the correlation between these two sets of grades. The diagram below shows the relations in the case of one factor in a population composed of $u\ AA + p\ Aa + v\ aa$. The coefficient of correlation (product-moment method) comes out:

$$r = \frac{v\,(2\,u + p)}{\sqrt{v\,(u + p)\,(up + vp + 4\,uv)}}$$

	$A-$	aa
AA	u	
Aa	p	
aa		v

It is easy to show that this formula applies to any number of factors in which the three classes are distributed as above, regardless of whether the dominants act in the same or opposite sense on development.

There are two cases which are of special interest in which the formula simplifies considerably. If the population is in equilibrium under random mating, $p = \sqrt{uv}$ and the correlation comes out $\sqrt{\frac{2v}{v + \sqrt{v}}}$. Table 1 gives examples of this case:

TABLE 1

GAMETES		POPULATION			CORRELATION BETWEEN ZYGOTIC FORMULAE WITH DOMINANCE AND WITHOUT DOMINANCE
A	a	AA	Aa	aa	
$\frac{9}{10}$	$\frac{1}{10}$	$\frac{81}{100}$	$\frac{18}{100}$	$\frac{1}{100}$	$\sqrt{\frac{2}{11}} = 0.426$
$\frac{3}{4}$	$\frac{1}{4}$	$\frac{9}{16}$	$\frac{6}{16}$	$\frac{1}{16}$	$\sqrt{\frac{2}{5}} = 0.632$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	$\sqrt{\frac{2}{3}} = 0.817$
$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{6}{16}$	$\frac{9}{16}$	$\sqrt{\frac{6}{7}} = 0.926$
$\frac{1}{10}$	$\frac{9}{10}$	$\frac{1}{100}$	$\frac{18}{100}$	$\frac{81}{100}$	$\sqrt{\frac{18}{19}} = 0.970$

By assuming that there are an equal number of dominant and recessive gametes, i.e., that $u = v$, the general formula becomes much simplified even without assuming that there is equilibrium, the correlation becoming $\sqrt{\frac{1}{1 + p}}$. If there is equilibrium, the value of p must be $\frac{1}{2}$ in this case, giving as before, $\sqrt{\frac{2}{3}}$.

RELATIONS BETWEEN ZYGOTE AND GAMETE

The genetic constitution (H) is determined completely by that of the egg and of the sperm which unite at fertilization. There is in general equal inheritance from the parents. Let a be the path coefficient from gamete to zygote. If there is assortative mating from any cause, there will be some correlation between the gametes which unite. Represent this correlation by f. By application of equation (1):

$$2\,a^2 + 2\,a^2 f = 1$$

$$a^2 = \frac{1}{2\,(1 + f)}$$

This equation does not of course apply in a case in which there is unequal inheritance from the sexes, as where sex linkage is involved. Caution must also be used in applying it where the mated individuals belong to different generations in a system of consanguine mating.

The constitution of the germ-cells is in part determined by that of the parent, but only in part, since the same parent can produce numerous different kinds of germ-cells. The other factor is chance at segregation. In order to find the value of the path coefficient from zygote to germ-cell, it is necessary to know the relation between the germ-cells produced by gametogenesis from a single gametocyte. Let this correlation be represented by g, and let

$$p_{G \cdot H''} = r_{GH''} = b$$

So far as one pair of allelomorphs is concerned, it is obvious that the relation between the products of gametogenesis must be the same as that between the egg and sperm which united to produce the parent. If the parents are a random sample from the preceding generation, an important assumption, we have: $g = f'$.

The prime in the symbol f' is used to indicate the preceding generation. The nomenclature f', f'', f''' will be used for the first, second and third generations preceding the one in question.

In the case of multiple factors, the same formula applies, provided that the factors of different sets of allelomorphs combine with each other at random at fertilization. This rules out assortative mating based on somatic resemblance. It does not rule out linkage in a population which is in equilibrium or has departed from equilibrium through consanguine matings. Assortative mating and selection (parents not a random sample of preceding generation) will be considered later.

We have considered the constitution of a germ-cell as in part determined by that of the parent and in part by chance at fertilization. In a sense, it is legitimate to reverse the arrows from parental constitution to germ-cell, and consider the former as completely determined by two gametes produced by gametogenesis. The path coefficient $p_{H''\cdot G}$ is, of course, not the same as that for the reverse direction $p_{G\cdot H''}$.

$$2\,p^2_{H''\cdot G} + 2\,p^2_{H''\cdot G}\,g = 1 \qquad \text{(from equation (1))}$$

$$r_{H''G} = p_{H''\cdot G}\,(1 + g) = b \qquad \text{(from equation (2))}$$

By elimination of $p_{H''\cdot G}$ we obtain

$$b^2 = \tfrac{1}{2}\,(1 + g)$$

The correlation between the egg and sperm depends on that between the parental formulae which we will represent by m.

$$f = b^2\,m$$

The correlation between the parents is greater or less than that between their genetic constitutions, depending on whether the assortative mating is based on somatic resemblance or consanguinity. In the former case, it is obviously $\dfrac{m}{h'^2}$ in the absence of dominance, $\dfrac{m}{h'^2}\,(1 + p)$ with perfect dominance and an equal number of dominant and recessive factors. In the case of consanguine mating the correlation between the parents is mh'^2 without dominance, and $\dfrac{mh'^2}{1 + p}$ with dominance and an equal number of dominant and recessive factors.

CORRELATIONS BETWEEN RELATIVES

Parent and offspring are connected by two chains of path coefficients, the direct path and the path through the other parent.

$$r_{po} = abhh' (1 + m) \quad \text{(no dominance)}$$

$$r_{po} = \frac{abhh'}{1+p} (1 + m) \quad \text{(typical case of dominance)}$$

Two brothers or sisters are connected by five chains of path coefficients, including common environmental influences.

$$\begin{aligned} r_{oo'} &= 2\,habbah + 2\,habmbah + e^2 \\ &= 2\,a^2b^2h^2 (1 + m) + e^2 \quad \text{(no dominance)} \end{aligned}$$

$$r_{oo'} = \frac{2\,a^2b^2h^2}{1+p} (1 + m) + e^2 \quad \text{(typical case of dominance)}$$

The correlation between any pair of relatives can easily be found by extension of this method.

THE DEGREE OF HOMOZYGOSIS AND HETEROZYGOSIS

In a population which is in equilibrium, under random mating, the percentage of homozygosis depends on the relative proportion of dominants and recessives. In a population composed of $x^2 AA + 2xy\,Aa + y^2\,aa$ the percentage of homozygosis is of course $\frac{x^2 + y^2}{(x + y)^2}$. This equals 50 percent only if there are an equal number of dominants and recessives.

Where there is not random mating, the percentage of homozygosis can be found from the correlation between gametes which unite at fertilization. Assume that there are x cases in which A unites with A, an equal number in which a unites with a, and $1 - 2x$ cases in which A unites with a. The product-moment correlation between the gametes, which we have called f, comes out $4x - 1$. Thus for the percentage of heterozygosis we have:

$$p = 1 - 2x = \tfrac{1}{2} (1 - f)$$

This formula applies to any number of factors in the absence of assortative mating based on somatic resemblance. The squared standard deviation for the variation due to a single Mendelian factor is obviously $1 - p$, and for n factors, combined at random, is $n(1 - p)$, where the effect of one factor is the unit. The genetic variation of the population thus increases with the percentage of homozygosis. As the variation due to other causes, not genetic, remains constant, the degree of determination by heredity, h^2, also increases unless there has been complete determination by heredity from the first.

$$h^2 = \frac{2\,h_0^2\,(1-p)}{h_0^2\,(1-2\,p)+1} \text{ where } h_0^2 \text{ is the initial value of } h^2$$

From the foregoing formulae, it is possible to express the correlations and path coefficients in each generation in terms of those in the preceding generation, provided that an expression can be found for m, the correlation between the genetic constitutions of the parents. This can easily be done for many systems of consanguine mating as will be brought out in a subsequent paper.

SUMMARY OF FORMULAE

The table below shows these equations, and the form which they take if it is assumed that the population is in equilibrium (m, g, f, a^2, b^2 constant) and in the case of random mating ($m = 0$).

TABLE 2

General formulae.

CONSANGUINE MATING	EQUILIBRIUM	RANDOM MATING
$h^2 + d^2 + e^2 = 1$	$h^2 + d^2 + e^2 = 1$	$h^2 + d^2 + e^2 = 1$
$m = \varphi\,(a'b'm')$	$m = \text{constant}$	$m = 0$
$g = f'$	$g = f' = f = \frac{m}{2-m}$	$g = f' = f = 0$
$b^2 = \frac{1}{2}\,(1+f')$	$b^2 = \frac{1}{2-m}$	$b^2 = \frac{1}{2}$
$f = b^2m$	$f = \frac{m}{2-m}$	$f = 0$
$a^2 = \frac{1}{2\,(1+f)}$	$a^2 = \frac{1}{4}\,(2-m)$	$a^2 = \frac{1}{2}$
$ab = \frac{1}{2}\sqrt{\frac{1+f'}{1+f}}$	$ab = \frac{1}{2}$	$ab = \frac{1}{2}$
$p = \frac{1}{2}\,(1-f)$	$p = \frac{1-m}{2-m}$	$p = \frac{1}{2}$
$h^2 = \frac{2\,h_0^2\,(1-p)}{h_0^2\,(1-2\,p)+1}$	$h^2 = \frac{2\,h_0^2}{2-m\,(1-h_0^2)}$	$h^2 = h^2$
$r_{pp} = mh'^2$	$r_{pp} = mh^2$	$r_{pp} = 0$
$r_{po} = abhh'\,(1+m)$	$r_{po} = \frac{1}{2}\,h^2\,(1+m)$	$r_{po} = \frac{1}{2}\,h^2$
$r_{oo} = 2\,a^2b^2h^2\,(1+m)+e^2$	$r_{oo} = \frac{1}{2}\,h^2\,(1+m)+e^2$	$r_{oo} = \frac{1}{2}\,h^2+e^2$

To allow for dominance (assuming equal number of dominant and recessive factors as in case of p) multiply the formulae for r_{pp}, r_{po} and r_{oo} by $\frac{1}{1+p}$.

It will be seen that if it is assumed that equilibrium is reached the formulae become very simple and still more so in the case of random mating. In this case, the degree of determination by heredity can be found directly from the correlation between parent and offspring. The correlation between parent and offspring and between brothers differs only in common influence of environmental factors on the latter. If these are assumed to be litter-mates for whom all or nearly all tangible environmental factors may be assumed to be in common, we can distinguish this class of factor from the chance variations in development which are not due to tangible external conditions.

ILLUSTRATION OF METHOD

A good illustration of these last points has been found in data on the inheritance of the piebald pattern in guinea-pigs, a case discussed in more detail elsewhere (WRIGHT 1920). The correlation between mated animals in a stock which had been bred at random was found to be + 0.019, indicating that there had been no unconscious assortative mating. The correlation between parent and offspring was +0.211, between litter mates +0.214.

$$\begin{aligned} r_{po} &= \tfrac{1}{2} h^2 = 0.211 & h^2 &= 0.422 \\ r_{oo} &= \tfrac{1}{2} h^2 + e^2 = 0.214 & e^2 &= 0.003 \\ & & d^2 &= \underline{0.575} \\ & & & \, 1.000 \end{aligned}$$

The standard deviation of the spotting in this stock was 0.802 in the units used. This means about 20 percent of the coat. The squared standard deviation can be broken into three parts, by multiplying by h^2, e^2, and d^2 giving the portions due respectively to heredity, tangible environment and chance in development. We find 0.271 due to heredity and 0.372 due to other causes. On inbreeding brother with sister for a dozen generations, one would expect to find the portion of the squared standard deviation due to heredity eliminated, leaving merely that due to other causes. Such a stock was actually on hand. Its standard deviation was found to be 0.603, giving a squared standard deviation of 0.364 in close agreement with expectation (0.372), (on the assumption that dominance is lacking, an assumption justified by the results of crosses between inbred families at opposite extremes in amount of white). The correlation between parent and offspring in the inbred family was only 0.014 and that between litter mates 0.069, indicating that heredity actually had been virtually eliminated by the inbreeding.

LITERATURE CITED

EAST, E. M., AND JONES, D. F., 1919 Inbreeding and outbreeding. 285 pp. Philadelphia: J. B. Lippincott Co.

FISH, H. D., 1914 On the progressive increase of homozygosis in brother-sister matings. Amer. Nat. **48**: 759–761.

HARDY, G. H., 1908 Mendelian proportions in a mixed population. Science **28**: 49–50.

JENNINGS, H. S., 1912 The production of pure homozygotic organisms from heterozygotes by self-fertilization. Amer. Nat. **45**: 487–491.

1914 Formulae for the results of inbreeding. Amer. Nat. **48**: 693–696.

1916 The numerical results of diverse systems of breeding. Genetics **1**: 53–89.

1917 The numerical results of diverse systems of breeding with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. Genetics **2**: 97–154.

PEARL, R., 1913 A contribution towards an analysis of the problem of inbreeding. Amer. Nat. **47**: 577–614.

1914 a On the results of inbreeding a Mendelian population; a correction and extension of previous conclusions. Amer. Nat. **48**: 57–62.

1914 b On a general formula for the constitution of the *n*th generation of a Mendelian population in which all matings are of brother × sister. Amer. Nat. **48**: 491–494.

PEARSON, K., 1904 On a generalized theory of alternative inheritance, with special reference to Mendel's laws. Phil. Trans. Roy. Soc. (A) **203**: 53–86.

ROBBINS, R. B., 1917 Some applications of mathematics to breeding problems. Genetics **2**: 489–504.

1918 a Applications of mathematics to breeding problems. II. Genetics **3**: 73–92.

1918 b Some applications of mathematics to breeding problems. III. Genetics **3**: 375–389.

1918 c Random mating with the exception of sister-by-brother mating. Genetics **3**: 390–396.

SUMNER, F. B., 1918 Continuous and discontinuous variations and their inheritance in Peromyscus. Amer. Nat. **52**: 177–208, 290–301, 439–454.

WENTWORTH, E. N., AND REMICK, B. L., 1916 Some breeding properties of the generalized Mendelian population. Genetics **1**: 608–616.

WRIGHT, S., 1920 The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. Proc. Nation. Acad. Sci. **6**: 320–332.

1921 Correlation and causation. Jour. Agric. Res. **20**: 557–585.