III. *Mathematical Contributions to the Theory of Evolution.*—XII. *On a Generalised Theory of Alternative Inheritance, with special Reference to* MENDEL'S *Laws.*

By KARL PEARSON, *F.R.S.*

Received September 11,—Read November 26, 1903.

(1.) *Introductory. On a Generalised Theory of Alternative Inheritance, with special Reference to* MENDEL'S *Laws.**

IT seems likely to be of interest at the present time to consider rather at length a fairly full mathematical theory of the pure gamete. We do not venture to call this theory a generalised Mendelian theory of inheritance, partly because it is not even the most general theory of the pure gamete conceivable, partly because MENDEL'S original theory of heredity was perfectly clear and perfectly simple, and is not the theory here developed. The pure and simple Mendelian theory seems to have been discarded in the light of recent experimental results by more than one Mendelian, both in this country and abroad. The original Mendelian theory has been replaced by what are termed "Mendelian Principles." In this aspect of investigation the fundamental principles propounded by MENDEL are given up, and for each individual case a pure gamete formula of one kind or another is suggested as describing the facts.† This formula is then emphasised, modified or discarded, according as it fits well, badly, or not at all with the growing mass of experimental data.

It is quite clear that it is impossible while this process is going on to term anything whatever Mendelian as far as theory is concerned. The present investigation is therefore not a generalised Mendelian theory of heredity: we speak of it merely as a generalised theory of alternative inheritance, and it is based on the conception that the gamete remains pure, and that the gametes of two groups, while they may link up to form a complete zygote, do not thereby absolutely fuse and lose their

* I owe the incentive to this memoir to Professor W. F. R. WELDON, who had already worked at some of the simpler special cases and who placed his results entirely at my disposal.

† See especially TSCHERMAK, 'Zeitschrift f. d. landwirthsch. Versuchswesen,' Jahrg. IV. ("Ueber Züchtung neuer Getreiderassen"); DE VRIES, 'Ber. d. deutsch. botan. Gesellsch.,' vol. xviii. (1900), pp. 435–443; BATESON, 'Proc. Camb. Phil. Soc.,' vol. 12, p. 53; 'Nature,' April and May, 1903.

The Royal Society is collaborating with JSTOR to digitize, preserve, and extend access to *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character.* www.jstor.org

identity. The analytical expression of this is represented by the fundamental formula:

$$\left.\begin{aligned}(\mathrm{AA}') \times (aa') \\ = (\mathrm{A} + \mathrm{A}')(a + a)\end{aligned}\right\} = \left\{\begin{aligned}&\mathrm{A}a \\ &\mathrm{A}a' \\ &\mathrm{A}'a \\ &\mathrm{A}'a'\end{aligned}\right. \quad \ldots\ldots\ldots \quad (\mathrm{i.}),$$

where (AA′) and (aa') are the parental zygotes, and the right-hand side of the equation represents the four possible constitutions of the offspring. Such a formula as the above may be accepted without any hypothesis as to dominant and recessive characters, but these terms were certainly essential to Mendelian theory as propounded by MENDEL himself, and it becomes very doubtful whether we ought to attach his name to any theory which discards these "recognition marks." It is very convenient, however, to have names for the alternative elements expressed by capital and small letters respectively. I propose for the purpose of this paper to term an A-element a *protogene*, and an a-element an *allogene*. Two protogenic elements will give rise to a protogenic zygote AA, two allogenic elements to an allogenic zygote aa, and a protogenic and allogenic element to what Mr. BATESON has termed a heterozygote Aa. We may thus class his homozygotes into protozygotes and allozygotes. We reach pure Mendelianism by making our protozygotes "dominants," our allozygotes, "recessives," and our heterozygotes "hybrids of dominant character." In so far as our theory of pure gametes replaces protozygote, allozygote, and heterozygote by "dominant," "recessive," and "hybrid with dominant character," it becomes a generalised Mendelian theory, but only in this case. Otherwise we must look upon it as an attempt—in one direction only of course—to give a consistent mathematical basis to the various formulæ which have been propounded for describing statistical data classed under Mendelian categories; shortly we shall endeavour to develop a general pure gamete theory.

The results were worked out in a purely impartial frame of mind; indeed, once state the hypotheses, and the analysis is far too complex to allow us to predict *à priori* what can possibly result from it, nor does the investigation admit of any but one solution. If the hypotheses are admissible, then any narrower pure gamete formula must lead to results embraced under our general conclusions.

What we have to admit at the present time are the following conditions:—

(i.) The existence of a vast bulk of evidence that heredity, as far as measurable characters are concerned, follows within a population perfectly definite laws.

(ii.) The existence of another mass of experiments, in which simple and pure Mendelianism is certainly inadmissible, but in which certain ratios undoubtedly approach the values they would have on such a simple and pure Mendelian theory.

It is possible, therefore, that a generalised theory of the pure gamete would account

for (ii.); it can only do so satisfactorily, however, if it does not contradict the results of (i.). Hence arises the present attempt to develop in one direction a generalised theory of the action of the germ-cell.

As we have frequently had to assert, the laws investigated under (i.) have nothing whatever to do with any physiological hypothesis. That a physiological hypothesis leads to them is not much test of its validity—it is a necessary, but not *sufficient*, criterion of its correctness. If, however, it contradicts them, we are bound to discard it, and seek for its modification or replacement. The present study is an attempt to see how far one generalised pure gamete theory leads to results in accordance with the law of regression and the known nature of the distributions of offspring in populations.

(2.) *Nature of Hypothesis adopted.*

We start with a zygote consisting not of a single protogenic pair AA, but built up of n such pairs,

$$A_1A_1 + A_2A_2 + A_3A_3 + \ldots + A_nA_n.$$

We suppose this to produce gametes which unite with those of a similar allogenic zygote

$$a_1a_1 + a_2a_2 + a_3a_3 + \ldots + a_na_n.$$

Any element of the protogenic gamete must unite with the corresponding element of the allogenic gamete, *i.e.*, A_r with a_r, and by the fundamental principle (i.) above, this gives rise to the four possibilities

$$A_ra_r,$$
$$A_ra_r,$$
$$a_rA_r,$$
$$a_rA_r,$$

which are all of the same constitution. The result is the hybrid group, symbolised by

$$a_1A_1 + a_2A_2 + a_3A_3 + \ldots + a_nA_n,$$

the perfect multiple heterozygote.

The population will now be supposed to consist of any number of such perfect heterozygotes, which we shall suppose to again cross. We shall now have

$$a_rA_r \times a_rA_r = \begin{cases} a_ra_r \\ a_rA_r \\ A_ra_r \\ A_rA_r, \end{cases}$$

or, each couplet will give rise to four possibilities, representing, however, only three constitutional differences, expressed by

$$a_r a_r + 2a_r \mathrm{A}_r + \mathrm{A}_r \mathrm{A}_r.$$

Since these four possibilities may occur with each of the n couplets, we shall have when two perfect heterozygotes cross, 4^n resulting possibilities. These form the resulting population of the second generation. Our first problem will be to find the distribution of this population. This, according to MENDEL, is the segregating generation. We must inquire into the frequency of each constitutional difference in this segregating generation.

We now reach our second limiting hypothesis, which is needful if we are to apply our theory to sexual reproduction. We suppose :—

There to be an absence of homogamy (including self-fertilisation), and the members of the second generation to cross absolutely at random and with equal fertility.* We have then to ask what is the distribution of constitutional differences in the third generation. Does the process of segregation begun in the second generation continue to the third, or does the population now remain stable? Is the *continual* segregation into pure protogenic and allogenic individuals a *necessary* result of any pure gamete theory, or does the belief in such necessity depend upon the first Mendelian experimenters working only with self-fertilising individuals?

(3.) PROBLEM I.—*To find the Distribution of the Offspring of the Perfect Heterozygotes.*

We shall here use a symbolic form of analysis. Let u stand for aa, v for $a\mathrm{A}$, and w for AA; then any corresponding couplets will give rise to

$$u + 2v + w,$$

and any one of these constitutions may be associated with one of the similar constitutions in any of the remaining $n - 1$ couplets. Hence the general distribution of the population will be given by the terms of the multinomial

$$(u + 2v + w)^n.$$

This equals†

$$u^n + nu^{n-1}(2v + w) + \frac{n(n-1)}{1 \cdot 2} u^{n-2}(2v + w)^2 + \ldots + c_{n,s,0}\, u^{n-s}(2v + w)^s + \ldots$$

* If one is to study heredity in populations with a view to the problem of evolution, the conditions as to fertilisation should approach as far as possible the conditions we suppose them to be under in a natural state; we must fix our attention on the *mass* relations between successive generations of the population.

† Throughout this memoir the symbol $c_{n,p,q}$ is used for the expression $\lfloor n \,/\, \{\lfloor n-p-q \rfloor p \rfloor q\}$.

Thus, for example, there would be out of the total population of possibilities 4^n: 1 purely allogenic individual, $n \times 3$ individuals with $n - 1$ allogenic couplets; $2n$ of these would have one heterogenic couplet, and n would have one protogenic couplet.

Generally there would be $3^s\ c_{n,s,o}$ individuals with $n - s$ allogenic couplets, and these individuals would be distributed according to the terms of the binomial $(2v + w)^s$.

We are thus able to write down at once the number of any class of individual that can appear in the segregating generation. For example, how often do individuals like $u^{n-p-q}v^p w^q$ appear, *i.e.*, individuals with $n - p - q$ allogenic, p heterogenic, and q protogenic couplets?

To answer this problem all we have to do is to pick out the coefficient of $u^{n-p-q}v^p w^q$ in the above multinomial, and the result is

$$2^p c_{n,p,q}.$$

We are thus fully able to predict how many individuals of each kind ought to occur when a population of perfect n-couplet heterozygotes are crossed.

Corollary (i.).—Let us consider only the number of allogenic couplets in the distribution of the segregating generation. If we were "pure Mendelians" we should for the purpose of character classification make $v = w$, as the heterogenic couplet would then give the dominant character. But without doing this we can assume v and w to be non-u's.

Hence the distribution of allogenic characters in the population follows the simple binomial

$$4^n \left(\tfrac{1}{4} + \tfrac{3}{4}\right)^n.$$

Thus we see that the distribution would be a skew binomial closely approximating to my skew curve of Type III.,* and becoming indefinitely close to a normal distribution of the form

$$y = y_0 e^{-x^2/2\sigma^2},$$

when the number of couplets, n, is indefinitely increased.

For any value of n the mean of the skew binomial as measured by the formula of the memoir on "Skew Variation,"† is

$$= n + 1 - \left(1 + \tfrac{3}{4}n\right) = \tfrac{1}{4}n,$$

and the standard deviation $= \sqrt{\tfrac{3}{16}n}$.

Thus the mean number of allogenic couplets in the members of the segregating generation is $\frac{1}{4}$ of the total number of couplets.

* 'Phil. Trans.,' A, vol. 186, p. 373.

† *Ibid.*, p. 346.

Corollary (ii.).—The distribution of heterogenic couplets in the segregating generation is given by the symmetrical binomial

$$4^n \left(\tfrac{2}{4} + \tfrac{2}{4}\right)^n.$$

The mean number is therefore $\frac{1}{2}n$, and the standard deviation $= \sqrt{\frac{1}{4}n}$.

This is a symmetrical binomial, and approaches extremely closely, even for a fairly small value of n, to the normal curve. We see that if any character depends solely upon heterogenic couplets, the distribution will be nearly normal, and the variability slightly greater than one depending on the allogenic (or, of course, the protogenic) couplets only.

To sum up, then, so far as the distribution of characters depending upon allogenic or heterogenic couplets goes, we may say that a generalised theory of the pure gamete leads us to those normal and skew distributions of frequency with which biometric studies of variability have made us already familiar. It would not be possible to base a crucial experiment on the existence or non-existence of such frequency distributions. The generalised pure gamete theory would, however, account for their appearance, which, of course, a purely descriptive statistical theory cannot do. On the other hand, distributions diverging much beyond the errors of random sampling from binomials of the above types would tell *pro tatno* against the pure gamete theory in its above form. The presence of binomials of two types only, $\left(\frac{1}{2} + \frac{1}{2}\right)^n$ and $\left(\frac{1}{4} + \frac{3}{4}\right)^n$, ought to be capable of detection, even if it would not already have been discovered, had it been the rule.

(4.) PROPOSITION II.—*To determine the Distribution of the Offspring of the Segregating Generation, supposing them to Mate at Random and without Differential Fertility.*

The solution of this problem may be reached as follows:—

Suppose P any male, and Q any female, say each of $n - 1$ couplets, producing an array of offspring, which we will denote by R; now suppose an additional couplet, the n^{th}, added to both male and female zygote. The male may be now:

$$\mathrm{P} + a_n a_n, \quad \text{or} \quad \mathrm{P} + a_n \mathrm{A}_n, \quad \text{or} \quad \mathrm{P} + \mathrm{A}_n a_n, \quad \text{or} \quad \mathrm{P} + \mathrm{A}_n \mathrm{A}_n;$$

and the female may be

$$\mathrm{Q} + a_n a_n, \quad \text{or} \quad \mathrm{Q} + a_n \mathrm{A}_n, \quad \text{or} \quad \mathrm{Q} + \mathrm{A}_n a_n, \quad \text{or} \quad \mathrm{Q} + \mathrm{A}_n \mathrm{A}_n;$$

that is, we get 4×4 new mating individuals, with $4 \times 4 \times 4$ new offspring possibilities.

Now consider the first father $\mathrm{P} + a_n a_n$; the possibilities which arise from mating him with the four mothers are the array R of offspring combined with any one of the

16 possibilities $8a_na_n + 8a_nA_n$, or this is the same thing as multiplying the R array by the symbolic factor $8(u + v) = 16U$, say.

The next pair of fathers $2(P + a_nA_n)$ with the four mothers reproduce the array R of offspring combined with $8(a_na_n + 2a_nA_n + A_nA_n)$, or 32 possibilities. But this is the same as multiplying the R-array by the symbolic factor $8(u + 2v + w) = 32V$, say. Lastly, the $P + A_nA_n$ father with the four mothers gives 16 possibilities of the form $8a_nA_n + 8A_nA_n$ to be combined with the R-array of offspring, which is the same thing as multiplying the R-array by the symbolic factor $8(v + w) = 16W$, say.

We have at once the symbolic relation among the operators:

$$U + 2V + W = u + 2v + w;$$

and, further, the important result that the array of offspring due to any pair P and Q of $\overline{n-1}$-couplet parents can be converted into the arrays of offspring due to the 16 pairs of parents formed by adding an additional couplet to P and Q, by multiplying that array by the symbolic factor

$$16U + 32V + 16W = 16(u + 2v + w).$$

We have thus by induction a means of finding the array of offspring due to a population of parents of n couplets from the series of arrays due to a population of $n - 1$ couplets. Since all the arrays are to be multiplied by the same symbolic factor, we can multiply their total by this factor. Or the distribution of offspring of $(n - 1)$-couplet parents being J, that of n-couplet parents

$$= 16(u + 2v + w)J = 4 \times 4 \times 4 . (\tfrac{1}{4}u + \tfrac{2}{4}v + \tfrac{1}{4}w)J.$$

Now consider parents of one couplet, their distribution is given by $aa + 2aA + AA$, and they are to mate with the same series, $aa + 2aA + AA$.

But

$$aa \times aa = 4aa,$$

$$2(aa \times 2aA) = 2(4aa + 4aA),$$

$$2(aa \times AA) = 2(4aA),$$

$$2aA \times 2aA = 4aa + 8aA + 4AA,$$

$$2(2aA \times AA) = 2(4aA + 4AA),$$

$$AA \times AA = 4AA,$$

$$\text{Total} = 16aa + 32aA + 16AA.$$

$$= 16(u + 2v + w) = 4 \times 4 \times 4(\tfrac{1}{4}u + \tfrac{2}{4}v + \tfrac{1}{4}w)$$

symbolically.

Hence, by the above proposition, the distribution of offspring of parents of two couplets is

$$4 \times 4 \times 4 \,.\, (\tfrac{1}{4}u + \tfrac{2}{4}v + \tfrac{1}{4}w) \times 4 \times 4 \times 4 \,.\, (\tfrac{1}{4}u + \tfrac{2}{4}v + \tfrac{1}{4}w)$$
$$= 4^2 \times 4^2 \times 4^2 \,.\, (\tfrac{1}{4}u + \tfrac{2}{4}v + \tfrac{1}{4}w)^2,$$

and, by induction, the distribution of offspring for the random mating of parents of n couplets is

$$4^n \times 4^n \times 4^n \,.\, (\tfrac{1}{4}u + \tfrac{2}{4}v + \tfrac{1}{4}w)^n.$$

This, except for the constant factor $4^n \times 4^n$, is absolutely identical with the distribution of the parental population, and accordingly if the next generation also mates at random, the mixed race will continue to reproduce itself without change. We therefore reach the following result :—

However many couplets we suppose the character under investigation to depend upon, the offspring of the hybrids—or the segregating generation—if they breed at random inter se, *will not segregate further, but continue to reproduce themselves in the same proportions as a stable population.*

It is thus clear that the apparent want of stability in a Mendelian population, the continued segregation and ultimate disappearance of the heterozygotes, is solely a result of self-fertilisation; with random cross fertilisation there is no disappearance of any class whatever in the offspring of the hybrids, but each class continues to be reproduced in the same proportions. Thus our generalised theory lends no countenance to the appearance of any "mutations" within a hybrid population under random mating; the only appearance of new constitutions is in the segregating generation, or the first generation of hybrid offspring. Except at this stage, the appearance of the unfamiliar is only the chance occurrence of a very rare normal variation. When we recollect that a purely allogenic individual is only to be expected once in a population of 4^n individuals, or if there be ten couplets, once in more than a million individuals, it will be clearly seen that the rarity of some of the more exceptional normal constitutions may easily lead to their being looked upon as "mutations," even if they appear in the offspring of a population many generations removed from hybridisation.

(5.) Proposition III.—*To find the Array of Offspring due to a Parent of given Gametic Constitution mating at Random.*

This can be again deduced by the method of induction adopted in the last proposition.

Supposing a male P of $n-1$ couplets to mate with all possible females, and R_{n-1} to be the array of offspring, then we have seen in the last proposition that if we add an n^{th} couplet $a_n a_n$ to P, the array of offspring due to $P + a_n a_n$ will be $16UR_{n-1}$; if we add a couplet $a_n A_n$, the array of offspring due to fathers of type $P + a_n A_n$ will be

$16\mathrm{VR}_{n-1}$, and if we add a couplet of form $\mathrm{A}_n\mathrm{A}_n$, the array will be of the form $8\mathrm{WR}_{n-1}$. Now start with a father of one couplet; this must be a_1a_1, or $a_1\mathrm{A}_1$, or $\mathrm{A}_1\mathrm{A}_1$, or in our symbolic notation u, v, or w; the offspring array are respectively $8a_1a_1 + 8a_1\mathrm{A}_1$ or $4a_1a_1 + 8a_1\mathrm{A}_1 + 4\mathrm{A}_1\mathrm{A}_1$, or $8a_1\mathrm{A}_1 + 8\mathrm{A}_1\mathrm{A}_1$, *i.e.*, 16U, 16V, or 16W. These, therefore, are the possible values of R_1. Hence, by the principle just developed above, the array of offspring due to a father of type

$$u^{n-p-q}\, v^p\, w^q$$

is

$$(16\mathrm{U})^{n-p-q}\,(16\mathrm{V})^p\,(16\mathrm{W})^q,$$

or remembering that such fathers occur with a frequency of $2^p c_{n,p\ q}$, we have for the total distribution of offspring of all fathers of type

$$u^{n-p-q}\, v^p\, w^q,$$

the symbolic result

$$4^n \times 4^n \,.\, c_{n,p,q}\mathrm{U}^{n-p-q}(2\mathrm{V})^p\mathrm{W}^q.$$

Substituting, the following expression would give all offspring of fathers of the type $u^{n-p-q}v^pw^q$, *i.e.*, with $n-p-q$ allogenic, p heterogenic, and q protogenic couplets

$$4^n \times 4^n \,.\, c_{n,p,q}\left(\tfrac{1}{2}u + \tfrac{1}{2}v\right)^{n-p-q}\left(\tfrac{1}{2}u + v + \tfrac{1}{2}w\right)^p\left(\tfrac{1}{2}v + \tfrac{1}{2}w\right)^q.$$

Therefore, given n and given p and q, it is merely a matter of expansion to find the array of offspring due to any special class of father.

Corollary (i.).—So far we have supposed our special class of father to be defined by the exact couplet distribution constitutional to him. But it is of interest to consider the array of offspring we get supposing only the allogenic couplets fixed in number, for example, in a generalised Mendelian theory if the number of recessive couplets be fixed, but the heterogenic and dominant, as both exhibiting dominant characters, be considered as indifferent. Let s = number of allogenic couplets, then we have to sum all arrays like

$$4^n \times 4^n \,.\, c_{n,p,q}\mathrm{U}^s(2\mathrm{V})^p\mathrm{W}^q,$$

subject to the condition that $p + q = n - s$.

The result is clearly

$$\begin{aligned}
&4^n \times 4^n \,.\, c_{n,s,o}\,\Sigma\left\{c_{p+q,p,o}\,(2\mathrm{V})^p\mathrm{W}^q\right\}\\
&\quad = 4^n \times 4^n \,.\, c_{n,s,o}\mathrm{U}^s\,(2\mathrm{V} + \mathrm{W})^{n-s}\\
&\quad = 4^n \times 4^n \,.\, c_{n,s,o}\left(\tfrac{1}{2}u + \tfrac{1}{2}v\right)^s\left(\tfrac{1}{2}u + \tfrac{3}{2}v + w\right)^{n-s}\\
&\quad = 4^n \times 4^n \,.\, c_{n,s,o}\left(\tfrac{1}{2}u + \tfrac{1}{2}v\right)^s\left\{\left(\tfrac{1}{2}u + \tfrac{1}{2}v\right) + (v + w)\right\}^{n-s}.
\end{aligned}$$

This, we note, is not a pure binomial, or the arrays of offspring of a father with a

given allogenic constitution are not either symmetrical or skew binomials, but of a much more complex character. The only exception is the array of offspring of pure allogenic fathers,* which is given by

$$4^n \times 4^n \times (\tfrac{1}{2}u + \tfrac{1}{2}v)^n.$$

This is a symmetrical binomial. This result is, of course, of special interest, for it gives us the distribution of offspring if the hybrid offspring were at any time crossed with the pure allogenic race, which was one of the original factors of the hybridisation. The deviation from binomial distribution in the above arrays ought to be further considered, for if this deviation should turn out to be very significant, it would form a convenient test for any generalised theory of pure gametes.

Corollary (ii.).—If we sum the above expressions for the array of offspring of all fathers of p allogenic couplets for values of s from o to n, we have the total offspring population

$$\begin{aligned} &= 4^n \times 4^n \,.\, \Sigma c_{n,\,s,\,o} \mathrm{U}^s (2\mathrm{V} + \mathrm{W})^{n-s} \\ &= 4^n \times 4^n \times (\mathrm{U} + 2\mathrm{V} + \mathrm{W})^n \\ &= 4^n \times 4^n \times (u + 2v + w)^n, \end{aligned}$$

a result we have already found in Proposition II. as giving the distribution of the total offspring population.

(6.) Proposition IV.—*To find the Mean Number of Allogenic Couplets in the Offspring of all Fathers having in their Constitution s-allogenic Couplets.*

By the first corollary to the last proposition the distribution of such offspring is given by

$$4^n \times 4^n \,.\, c_{n,\,s,\,o} (\tfrac{1}{2}u + \tfrac{1}{2}v)^s \{\tfrac{1}{2}u + \tfrac{1}{2}v + 2\eta\}^{n-s},$$

where η is written for $\frac{1}{2}(v + w)$, a quantity which is unity so long as we consider not the distribution, but the total number of the non-allogenic couplets. Now this is clearly the sum of a number of symmetrical binomials in $\frac{1}{2}u + \frac{1}{2}v$, and may be put

$$= 4^n \times 4^n \,.\, \sum_{i=o}^{i=n-s} c_{n,\,s,\,i} (\tfrac{1}{2}u + \tfrac{1}{2}v)^{n-i} (2\eta)^i.$$

Now the means of each of these binomials can be found from the general theory of the binomial.† If we take our origin at $\overline{n+1}$ allogenic couplets, with a frequency zero, the mean of the first binomial, or

* Or, of course, the array of sons from pure protogenic fathers.

† 'Phil. Trans.,' A, vol. 186, p. 373.

$(\frac{1}{2}u+\frac{1}{2}v)^n$ is at $1+\frac{1}{2}n$, and its total frequency $f_1 = 4^n \times 4^n \times c_{n,s,o}$;

the mean of the second binomial, or

$(\frac{1}{2}u+\frac{1}{2}v)^{n-1}$ . . $2+\frac{1}{2}(n-1)$. . . . . $f_2 = \frac{n-s}{1} 2f_1$;

the mean of the third binomial, or

$(\frac{1}{2}u+\frac{1}{2}v)^{n-2}$ . . $3+\frac{1}{2}(n-2)$. . . . . $f_3 = \frac{(n-s)(n-s-1)}{1.2} 2^2 f_1$;

the mean of the $(i+1)^{\text{th}}$ binomial

$(\frac{1}{2}u+\frac{1}{2}v)^{n-i}$ . . $i+1+\frac{1}{2}(n-i)$. $f_{i+1} = \frac{(n-s)(n-s-1)\ldots(n-s-i+1)}{1.2.3\ldots i} 2^i f_1$.

The total frequency is accordingly

$$\begin{aligned} F_s = f_1 + f_2 + f_3 + \ldots &= f_1 (1+2)^{n-3} \\ &= 4^n \times 4^n c_{n,s,o} 3^{n-s}. \end{aligned}$$

Hence if m_s be the distance from the same origin of the mean of the above system of binomials

$$\begin{aligned} f_1 \times 3^{n-s} \times m_s &= f_1\left(1+\frac{n}{2}\right) + f_2\left(2+\frac{n-1}{2}\right) + \ldots + f_{i+1}\left(i+1+\frac{n-i}{2}\right) + \ldots \\ &= f_1 \left\{ 1 + \frac{n}{2} + 2(n-s)\left(2+\frac{n-1}{2}\right) + 2^2 \frac{(n-s)(n-s-1)}{1.2}\left(3+\frac{n-2}{2}\right) \right. \\ &\quad \left. + \ldots + 2^i \frac{(n-s)(n-s-1)\ldots(n-s-i+1)}{1.2.3\ldots i}\left(i+1+\frac{n-i}{2}\right) + \ldots \right\}. \end{aligned}$$

Now

$$(1+2x)^{n-s} = \left\{ 1 + 2(n-s)x + 2^2 \frac{(n-s)(n-s-1)}{1.2} x^2 + \ldots \right\}.$$

Multiply by x^2, differentiate both sides and divide by 2, finally putting $x = 1$, and we find

$$\begin{aligned} 3^{n-s} + (n-s)3^{n-s-1} &= 1 + 2(n-s)\tfrac{3}{2} + 2^2 \frac{(n-s)(n-s-1)}{1.2}\tfrac{4}{2} + \ldots \\ &\quad + 2^i \frac{(n-s)(n-s-1)\ldots(n-s-i+1)}{1.2.3\ldots i} \frac{i+2}{2} + \text{\&c.} \end{aligned}$$

Hence we deduce

$$f_1 \times 3^{n-s} m_s = f_1 \{ 3^{n-s} + (n-s)3^{n-s-1} + \tfrac{1}{2}n 3^{n-s} \},$$

or

$$m_s = 1 + \tfrac{1}{2}n + \tfrac{1}{3}(n-s).$$

But the mean of the whole population of offspring is at $1 + \frac{3}{4}n$ from our origin. Thus we have the final results:

Mean number of allogenic couplets in offspring of fathers with s allogenic couplets

$$= \tfrac{1}{2}n - \tfrac{1}{3}(n - s) \text{ allogenic couplets.}$$

Deviation from mean of general population of this array of offspring

$$= \tfrac{1}{4}n - \tfrac{1}{3}(n - s) = \tfrac{1}{12}(4s - n).$$

Deviation of fathers from mean of population

$$= s - \tfrac{1}{4}n = \tfrac{1}{4}(4s - n).$$

Thus

$$\frac{\text{Deviation of offspring from mean of population}}{\text{Deviation of fathers from mean of population}} = \tfrac{1}{3}.$$

We have then the following results, which could certainly not have been foreseen:—

(a.) The regression is constant for all arrays, or the regression curve is a straight line.

(b.) The slope of this straight line is $\frac{1}{3}$, or, since we have seen that the population is stable, the parental correlation is $\frac{1}{3}$ also.

Now these results are of very singular importance. A very general theory of the pure gamete type leads to linearity of the regression curve, a result amply verified by observations on inheritance in populations;* and this result is quite independent of the number of couplets supposed to form the total character of the parent, or of the fact that in this case the arrays of offspring are skew and do not obey the normal law.† Further, the value of the correlation reached is numerically identical with the value obtained by FRANCIS GALTON in his original investigations on the inheritance of stature! The generalised theory of the pure gamete is thus shown, whatever the number of couplets taken, to lead to precisely the chief results already obtained by those who have studied heredity statistically. So far then it might appear that a generalised theory of the pure gamete was capable of being brought into accordance with the chief results of biometric experience in heredity. This would undoubtedly be a great step forward, as linking up perfectly definite inheritance results with a physiological theory of heredity. Unfortunately the whole drift of recent biometric observations on heredity emphasises three points:

First.—That the parental correlation appears to be markedly greater than $\frac{1}{3}$, nearer to ·45 to ·5.

* GALTON, 'Natural Inheritance,' p. 96; 'Biometrika,' vol. 2, pp. 216 and 362–3.

† This is further demonstration that linearity of regression has nothing whatever to do with the Gauss-Laplacian law of errors, *i.e.*, normal curves or surfaces.

Secondly.—That this correlation appears to vary slightly from character to character.

Thirdly.—That it does not appear to be absolutely the same for all species.

It is most unfortunate for this general theory of the pure gamete, that it throws the Mendelian back into the position of the biometrician of 1885.* One might have hoped that the generality involved in n couplets would have led to the requisite elasticity, or, failing this, to a numerical value of parental correlation nearer the cluster point of existing measurements than $\frac{1}{3}$. We can only say, at present, that a generalised theory of the pure gamete leads to precisely the same general features of regression as have been observed by the biometricians, but it appears numerically too narrow to describe the observed facts.

(7.) Proposition V.—*To find the Standard Deviation of the Array of Offspring due to Fathers with s-allogenic Couplets.*

We have to find the standard deviation σ_s of the combination of binomials dealt with in the previous proposition. Each component standard deviation must, of course, be weighted with the total frequency of the component, and there must be the proper reduction to the mean of the array as a whole.

The $(i+1)^{\text{th}}$ binomial $(\frac{1}{2}u + \frac{1}{2}v)^{n-i}$ has $\sqrt{(n-i)\tfrac{1}{2}\times\tfrac{1}{2}}$† for its standard deviation, and the distance of its mean from the mean of the array

$$= \{n+1-(i+1+\tfrac{1}{2}(n-i))\} - \{\tfrac{1}{2}n - \tfrac{1}{3}(n-s)\}$$
$$= \tfrac{1}{3}(n-s) - \tfrac{1}{2}i.$$

Further, the frequency of this component is

$$= c_{n-s,\,i,\,o}\, 2^i f_1.$$

We thus see that it contributes

$$c_{n-s,\,i,\,o}\, 2^i f_1 \left\{\frac{n-i}{4} + \left(\frac{n-s}{3} - \frac{i}{2}\right)^2\right\}$$

to the total second moment about the mean of the array. This gives us

$$f_1 \times 3^{n-s} \times \sigma_s^2 = f_1 \Sigma c_{n-s,\,i,\,o}\, 2^i \left\{\frac{n-i}{4} + \left(\frac{n-s}{3} - \frac{i}{2}\right)^2\right\},$$

* "Galton's law makes the amount of inheritance an absolute constant for each pair of relatives. It would thus appear not to be a character of race or species, or one capable of modification by natural selection." More ample statistical experience of populations since 1885 shows that *absolute* constancy of the heredity coefficients is not consonant with actual measurements.—'Roy. Soc. Proc.,' vol. 62, p. 411.

† 'Phil. Trans.,' A, vol. 185, p. 373.

or

$$3^{n-s}\sigma_s^2 = \Sigma c_{n-s,\,i,\,o}\, 2^i \left\{\frac{n}{4} + \frac{(n-s)^2}{9} + \frac{i\,(i-1)}{4} - i\frac{n-s}{3}\right\}.$$

Now

$$(1+2x)^{n-s} = \Sigma c_{n-s,\,i,\,o}\,(2x)^i,$$

and by differentiating

$$2\,(n-s)\,(1+2x)^{n-s-1} = \Sigma c_{n-s,\,i,\,o}\, 2^i i x^{i-1}.$$

Repeating the process

$$4\,(n-s)\,(n-s-1)\,(1+2x)^{n-s-2} = \Sigma c_{n-s,\,i,\,o}\, 2^i i\,(i-1)\,x^{i-2}.$$

Hence, putting $x = 1$, we have the required expressions on the right of the above result, or

$$3^{n-s}\sigma_s^2 = 3^{n-s}\left\{\frac{n}{4} + \frac{(n-s)^2}{9} + \frac{(n-s)\,(n-s-1)}{9} - \frac{2\,(n-s)^2}{9}\right\}.$$

Therefore

$$\sigma_s^2 = \tfrac{1}{4}n - \tfrac{1}{9}\,(n-s) = \tfrac{1}{36}\,(5n+4s).$$

Now the standard deviation of the whole population, as far as allogenic units are concerned, is

$$\sigma = \sqrt{n\tfrac{1}{4}\tfrac{3}{4}} = \sqrt{\tfrac{3}{16}n}.$$

Thus

$$\frac{\sigma_s}{\sigma} = \sqrt{\tfrac{4}{27}\left(5 + 4\,\frac{s}{n}\right)}.$$

This result is of singular interest. The variability of an array of offspring corresponding to a father of given allogenic constitution is not independent of the father, but increases steadily from a minimum of $\sqrt{\frac{5}{36}n}$, when there are no allogenic couplets in the father, to a maximum of $\sqrt{\frac{1}{4}n}$ when there are only allogenic couplets. In other words, fixing our attention on the same character, let the offspring of the hybrids *inter se* be crossed first with one pure race, and then with the second pure race, *i.e.*, first with pure allogenic and then with pure protogenic individuals, there ought to be a marked difference in the variability of the resulting offspring in the two cases.

Corollary (i.).—In the theory of *linear* regression as apart from the theory of normal correlation on the basis of the Gauss-Laplacian distribution,* if σ be the standard deviation of any character, and r its correlation with a second character, then

$$\sigma\sqrt{1-r^2}$$

* YULE, 'Journal of Royal Statistical Society,' vol. 60, December, 1897.

is the *mean* standard deviation of all the arrays of the first character for a given value of the second. This expression is no longer the actual standard deviation of each array.

It is of interest to see that this general law of linear regression is verified in the present case. We have $\sigma = \sqrt{\frac{3}{16}n}$ and $r = \frac{1}{3}$. Hence if Σ_m be the mean standard deviation of the arrays, we should expect

$$\Sigma_m^2 = \tfrac{3}{16}n\,(1 - \tfrac{1}{9}) = \tfrac{1}{6}n.$$

Remembering the weight of each array,* we have

$$4^n \times 4^n \times 4^n \times \Sigma_m^2 = \sum_{s=o}^{s=n} \left\{ 4^n \times 4^n c_{n,\,s,\,o} 3^{n-s} \frac{5n + 4s}{36} \right\}$$

$$= 4^n \times 4^n (1 + 3)^n \tfrac{5}{36}n + 4^n \times 4^n n 4^{n-1} \tfrac{4}{36}$$

$$= 4^n \times 4^n \times 4^n \times \tfrac{1}{6}n,$$

whence $\Sigma_m^2 = \frac{1}{6}n$, as we anticipated.

Corollary (ii.).—It is clear that some arrays of offspring will be more, others less variable than the general population. The standard deviation of an array will be equal to that of the general population when s is found from

$$\tfrac{1}{36}(5n + 4s) = \tfrac{3}{16}n, \qquad \text{or} \qquad s = \tfrac{7}{16}n.$$

Now the mean number of allogenic couplets in the general population is $\frac{1}{4}n$. Thus the offspring array equally variable with the general population is at distance $\frac{3}{16}n$ from the mean. But σ for the general population $= \sqrt{\frac{3}{16}n}$. Hence, if we take fathers deviating from the population mean by $\sqrt{\frac{3}{16}n} \times \sigma$, we should expect their offspring to be equally variable with the general population. Supposing, therefore, the theory under discussion were true, we should have a means of finding, at least approximately, the number n of couplets corresponding to the character under consideration. All we should have to do would be to find the standard deviation of each array of offspring corresponding to a given father; these standard deviations ought to increase or decrease steadily across the table, their squares giving a straight line when plotted. Smooth the results, interpolate a value equal to σ, and we shall have the character of the father whose offspring are equally variable with the general population; but the deviation of this father from the mean ought to be $\sqrt{\frac{3}{16}n} \times \sigma$. Hence, since σ can be found, we have at once an approximate value of n. *Approximate* only, of course, because our arrays are classed by units of measurement, inches,

* The number of offspring in the array due to the fathers with s-allogenic couplets is found at once by putting $u = v = w$ in the formula of Corollary (i.) on p. 61, and equals $4^n \times 4^n \cdot c_{n,\,s,\,o} 3^{n-s}$.

centimetres, &c., and each such unit will not, as a rule, represent one allogenic couplet; but interpolation ought to give a result not widely divergent from the truth.

The method would of course fail practically if n were very large. For example, if n were 48, the deviation of the required group of fathers would be 3σ, and hence such a father would only occur once in 1000 individuals. In a manageable population, therefore, we are very unlikely to have enough such fathers to form any reliable measure of the variability of their offspring. At the same time, the squares of the variabilities of the arrays of sons due to quite frequent fathers ought to give a straight line, and if this line be determined properly, there should be no difficulty in finding the theoretical position of the above father, and so finding n.

Many other physiological theories besides the present might give this peculiarity of the diminishing variability of the arrays of offspring as we pass from one side of the correlation table to the other. Such changes in variability are familiar to those who have had to deal with skew correlation. But, as far as we are aware, they have not hitherto been noticed in inheritance tables. The existence of this changing variability would not affect in any way the general theory of linear regression applied to heredity in populations. It would, however, lead to an immediate extension of that theory consisting in the tabling of the standard deviations of the arrays. Should the standard deviations of these arrays show no bias towards a linear distribution, but only the fluctuation to be expected from random sampling about the mean value $\sigma\sqrt{(1-r^2)}$, we should have a strong argument against the present general theory of alternative inheritance. We seem here, therefore, to have a crucial test of the validity of the theory, which may be quite as easy to apply as the previous test of the numerical value of the parental correlation.

Of course the results now reached are not consistent theoretically with normal correlation surfaces with their elliptic contour lines. The fact that Mr. GALTON came to his elliptic contours in the first instance on the basis of his observations, and not from any theory,* shows that they must in the case he was dealing with be approximately correct. Further, there is no doubt that in other statistics for characters in man there is within the limits of random sampling a close approximation to normal distribution. It might be hard to consider that such a deviation as would arise with a continuously increasing variability of the arrays from one side to the other of the table could exist and escape notice, had we not in physics had evidence that theory has often led to the discovery of an obvious relation which time after time must have been overlooked by previous observers unprovided with the theoretical hint of what to seek for. Hence while we may say that the parental correlation given by the theory is too rigid for the facts, we must leave this second test until more careful examination *ad hoc* has been made of the ample existing data.

* 'Natural Inheritance,' p. 101.

(8.) PROPOSITION VI.—*To find the Array of Offspring due to a Grandfather of s-allogenic Couplets, supposing Complete Random Mating in the Population.*

The general distribution of the population is

$$(u + 2v + w)^n.$$

The fathers with s-allogenic units in their correlation are given by

$$c_{n,\,s,\,o}\, u^s\, (2v + w)^{n-s}.$$

The array of offspring due to these fathers is simply obtained by writing U for u, V for v, and W for w, and multiplying by $4^n \times 4^n$. This is a general rule for getting the offspring from any father if he mates at random. It gives us, as on p. 61, for the offspring distribution

$$4^n \times 4^n \times c_{n,\,s,\,o}\, \mathrm{U}^s\, (2\mathrm{V} + \mathrm{W})^{n-s}$$

$$= 4^n \times 4^n \times c_{n,\,s,\,o}\, (\tfrac{1}{2}u + \tfrac{1}{2}v)^s\, \{(\tfrac{1}{2}u + \tfrac{1}{2}v) + (v + w)\}^{n-s}.$$

To get the offspring of this array treated as fathers and mating at random, we have only to repeat the process, and we find

Offspring of grandfather of s-allogenic couplets

$$= 4^n \times 4^n \times 4^n \times 4^n c_{n,\,s,\,o}\, (\tfrac{1}{2}\mathrm{U} + \tfrac{1}{2}\mathrm{V})^s\, \{\tfrac{1}{2}\mathrm{U} + \tfrac{3}{2}\mathrm{V} + \mathrm{W}\}^{n-s}$$

$$= 4^{4n} c_{n,\,s,\,o}\, (\tfrac{3}{8}u + \tfrac{1}{2}v + \tfrac{1}{8}w)^s\, (\tfrac{5}{8}u + \tfrac{1}{8}\tfrac{2}{8}v + \tfrac{7}{8}w)^{n-s}$$

$$= 4^{4n} c_{n,\,s,\,o}\, (\tfrac{3}{8}u + \tfrac{5}{8}\epsilon)^s\, (\tfrac{3}{8}u + \tfrac{5}{8}\epsilon + \tfrac{2}{5}\,(v + w))^{n-s} \times (\tfrac{5}{3})^{n-s},$$

where

$$\epsilon = \tfrac{4}{5}v + \tfrac{1}{5}w,$$

and is equal to unity if we identify v and w as something not allogenic. This can be dealt with exactly as in Proposition IV. we dealt with the array of offspring due to a father of s-allogenic couplets, *i.e.*, by analysing the array into the sum of a number of weighted binomials; in this case all skew.

Writing as before, $\eta = \frac{1}{2}\,(v + w)$, we have to expand

$$(\tfrac{3}{8}u + \tfrac{5}{8}\epsilon)^s\, (\tfrac{3}{8}u + \tfrac{5}{8}\epsilon + \tfrac{4}{5}\eta)^{n-s}.$$

The general term is

$$c_{n-s,\,i,\,o}\, (\tfrac{4}{5})^i \eta^i\, (\tfrac{3}{8}u + \tfrac{5}{8}\epsilon)^{n-i}.$$

This has a total frequency $c_{n-s,\,i,\,o}\, (\frac{4}{5})^i \times f_1$, and its mean is at a distance $i + 1 + \frac{5}{8}(n - i)$ from the origin which is taken at $(n + 1)$ allogenic couplets.

The total frequency of the array is $(1+\frac{4}{5})^{n-s}f_1$. Hence, if m'_s be the mean of the grandchildren measured from the same origin, we have

$$f_1 \times (\tfrac{9}{5})^{n-s} \times m'_s = f_1 \{1 + \tfrac{5}{8}n + \tfrac{4}{5}(n-s)\{2 + \tfrac{5}{8}(n-1)\}$$
$$+ (\tfrac{4}{5})^2 \frac{(n-s)(n-s-1)}{1.2}\{3 + \tfrac{5}{8}(n-2)\} \ldots$$
$$+ (\tfrac{4}{5})^i \frac{(n-s)(n-s-1)\ldots(n-s-i+1)}{1.2.3\ldots i}(i+1+\tfrac{5}{8}(n-i)\}$$
$$= f_1 \{(\tfrac{5}{8}n+1) \times (\tfrac{9}{5})^{n-s} + \tfrac{4\cdot 3}{5\cdot 8}(n-s)(\tfrac{9}{5})^{n-s-1}\},$$

or

$$m'_s = 1 + \tfrac{5}{8}n + \tfrac{4\cdot 3\cdot 5}{5\cdot 8\cdot 9}(n-s) = 1 + \tfrac{5}{8}n + \tfrac{1}{6}(n-s).$$

Thus

Mean of grandchildren $= \frac{3}{8}n - \frac{1}{6}(n-s)$.

Deviation from general population mean $= \frac{1}{8}n - \frac{1}{6}(n-s) = \frac{1}{24}(4s-n)$.

Deviation of grandparent from general population mean $= s - \frac{1}{4}n = \frac{1}{4}(4s-n)$.

Hence

$$\frac{\text{Deviation of offspring}}{\text{Deviation of grandparent}} = \tfrac{1}{6}.$$

This ratio is the same whatever be the allogenic constitution of the grandparent.

(9.) Proposition VII.—*To find the Array of Offspring due to an m^{th} Great-grandfather of s-allogenic Couplets, supposing Complete Random Mating in each Generation.*

The array due to a father of s-allogenic couplets is

$$4^n \times 4^n \times c_{n,s,o}\{\tfrac{1}{2}(u+v)\}^s\{\tfrac{1}{2}(u+v)+(v+w)\}^{n-s},$$

and, as we have already seen, we must multiply by $4^n \times 4^n$ and put $\frac{1}{2}(u+v)$ for u, $\frac{1}{4}(u+2v+w)$ for v, and $\frac{1}{2}(v+w)$ for w to get the array due to the grandparent of s allogenic units. This process must be repeated m times if we wish to obtain the array due to the m^{th} great-grandparent.

We must first investigate what happens to $\frac{1}{2}(u+v)$ if this interchange be made m times. Suppose that it has been done i times, and let the answer be

$$\mathrm{M}_i\tfrac{1}{2}(u+v) + \mathrm{M}'_i\tfrac{1}{2}(v+w).$$

Repeat the operation, and the expression becomes

$$(\tfrac{3}{4}\mathrm{M}_i + \tfrac{1}{4}\mathrm{M}'_i)\tfrac{1}{2}(u+v) + (\tfrac{1}{4}\mathrm{M}_i + \tfrac{3}{4}\mathrm{M}'_i)\tfrac{1}{2}(v+w),$$

or

$$M_{i+1} = \tfrac{3}{4}M_i + \tfrac{1}{4}M'_i,$$

$$M'_{i+1} = \tfrac{1}{4}M_i + \tfrac{3}{4}M'_i.$$

Therefore

$$M_{i+1} + M'_{i+1} = M_i + M'_i = M_0 + M'_0 = 1 + 0 = 1.$$

Hence

$$M_{i+1} = \tfrac{1}{2}M_i + \tfrac{1}{4}, \quad M'_{i+1} = \tfrac{1}{2}M'_i + \tfrac{1}{4},$$

$$M_{i+1} - \tfrac{1}{2} = \tfrac{1}{2}(M_i - \tfrac{1}{2}) = \frac{1}{2^{i+1}}(M_0 - \tfrac{1}{2}) = \frac{1}{2^{i+2}},$$

$$M'_{i+1} - \tfrac{1}{2} = \tfrac{1}{2}(M'_i - \tfrac{1}{2}) = \frac{1}{2^{i+1}}(M'_0 - \tfrac{1}{2}) = -\frac{1}{2^{i+2}}.$$

Hence, finally,

$$M_i = \tfrac{1}{2} + \frac{1}{2^{i+1}}, \quad M'_i = \tfrac{1}{2} - \frac{1}{2^{i+1}}.$$

Thus the result of m changes on $\frac{1}{2}(u + v)$ is known.

Similarly the result of m changes on $\frac{1}{2}(v + w)$ is

$$M_i \frac{(v + w)}{2} + M'_i \frac{u + v}{2}.$$

We can now write down the array of offspring due to an m^{th} great-grandparent of s-allogenic couplets. It is

$$(4^n \times 4^n)^m c_{n,s,o} \left(M_m \frac{u + v}{2} + M'_m \frac{v + w}{2}\right)^s \times \left\{(M_m + 2M'_m)\frac{u + v}{2} + (M'_m + 2M_m)\frac{v + w}{2}\right\}^{n-s}.$$

We must now find the mean of this array. For brevity let us write

$$M_m \tfrac{1}{2}(u + v) + M'_m \tfrac{1}{2}(v + w) = \mu u + \lambda\epsilon$$

$$(M_m + 2M'_m)\frac{u + v}{2} + (M'_m + 2M_m)\frac{v + w}{2} = \frac{M_m + 2M'_m}{M_m}(\mu u + \lambda\epsilon + \nu\eta),$$

where

$$\mu = \tfrac{1}{2}M_m, \quad \lambda = 1 - \tfrac{1}{2}M_m,$$

$$\epsilon = \frac{1}{1 + M'_m} v + \frac{M'_m}{1 + M'_m} w,$$

$$\nu = \frac{2(M_m^2 - M'^2_m)}{M_m + 2M'_m} = \frac{2(M_m - M'_m)}{1 + M'_m},$$

$$\eta = \tfrac{1}{2}(v + w)$$

Hence we have to find the mean of the system

$$(\mu u + \lambda\epsilon)^s (\mu u + \lambda\epsilon + \nu\eta)^{n-s}.$$

The i^{th} component binomial of this sum of binomials is

$$c_{n-s,i,o}\, \nu^i \eta^i (\mu u + \lambda\epsilon)^{n-i}.$$

It therefore has its mean at a distance

$$i + 1 + \lambda(n - i)$$

from $(n + 1)$ allogenic couplets, and a frequency given by

$$f_i = c_{n-s,i,o}\, \nu^i f_1.$$

The total frequency of the whole array $= (1 + \nu)^{n-s} f_1$.

Hence, taking moments round the origin at $n + 1$ allogenic couplets, we have, if m'_s be the mean of the array,

$$\begin{aligned}(1 + \nu)^{n-s} f_1 \times m'_s = f_1 \Big\{ & 1 + \lambda n + \nu(n - s)(2 + \lambda(n - 1)) \\ & + \nu^2 \frac{(n - s)(n - s - 1)}{1 \cdot 2} (3 + \lambda(n - 2)) + \ldots \\ & + \nu^i \frac{(n - s)(n - s - 1)\ldots(n - s - i + 1)}{1 \cdot 2 \cdot 3 \ldots i} (1 + i + \lambda(n - i)) + \ldots \Big\}.\end{aligned}$$

Summing and dividing by $(1 + \nu)^{n-s}$ we find

$$m'_s = 1 + \lambda n + \frac{(n - s)\nu(1 - \lambda)}{1 + \nu}.$$

Hence the mean number of allogenic couplets in the members of the array

$$= n + 1 - m'_s = n(1 - \lambda) - (n - s)\frac{\nu}{1 + \nu}(1 - \lambda).$$

Deviation of offspring from mean of general population

$$= s\frac{\nu(1 - \lambda)}{1 + \nu} - n\left\{\tfrac{1}{4} - \frac{1 - \lambda}{1 + \nu}\right\}.$$

We now substitute for ν and λ in terms of M_m and M'_m, and find

$$\frac{\nu(1 - \lambda)}{1 + \nu} = \tfrac{1}{3}(\mathrm{M}_m - \mathrm{M}'_m) = \tfrac{1}{3}\frac{1}{2^m}$$

$$\tfrac{1}{4} - \frac{1 - \lambda}{1 + \nu} = \tfrac{1}{4} - \tfrac{1}{6}(1 + \mathrm{M}'_m) = \frac{1 - 2\mathrm{M}'_m}{12} = \tfrac{1}{12}\frac{1}{2^m}.$$

Hence: Deviation of offspring $= \frac{1}{12}\frac{4s - n}{2^m}$; but the deviation of m^{th} great-grandparent $= s - \frac{1}{4}n$.

Thus we have

$$\frac{\text{Deviation of offspring}}{\text{Deviation of } m^{\text{th}} \text{ great-grandparent}} = \frac{1}{3}\frac{1}{2^m}.$$

This result is independent of s and of n.

Thus we conclude:

(i.) The regression of offspring on any individual ancestor is linear;

(ii.) The correlation coefficient is halved at each stage in ancestry;

(iii.) The result is perfectly independent of the number of couplets introduced into the formula.

The first two results are very familiar to biometric workers in heredity.

The actual numerical values of the grandparental, great-grandparental, great-great-grandparental correlations are $\frac{1}{6}$, $\frac{1}{12}$, $\frac{1}{24}$, &c.

These are distinctly less than the values so far reached for ancestral correlation, the grandparental correlations, for instance, lying between ·2 and ·3.

The results show, however, that a general theory of the pure gamete, embracing the simpler forms of the Mendelian principle, leads us directly to a series of ancestral correlations decreasing in a geometrical progression. Thus, when we suppose a population arising from hybridisation to cross at random, we find that it obeys the second fundamental assumption of the biometric theory of heredity.* In other words, ancestry is of the utmost importance, and the population follows laws identical in form with those propounded in the biometrical theory on the basis of a linear regression multiple correlation. Only the values of the constants deduced for the law of ancestral heredity from the present theory of the pure gamete (which appears to cover the bulk of Mendelian formulæ hitherto propounded) are sensibly too small to satisfy the best recent observations on inheritance.

It is of interest to find "Mendelian Principles" when given a wide analytical expression leading up to the very laws of linear regression, of distribution of frequency, and of ancestral inheritance in populations, which have been called into question as exhibiting only a blurred and confused picture of what actually takes place.

It would be an immense advantage if we could accept such a theory of the pure gamete as has been here analysed as a physiological basis for the theory of heredity. We should then have a physiological origin for the ideas of regression and of ancestral inheritance which statistics of heredity in populations have made familiar to biometric workers. Unfortunately, even such a general pure gamete theory as we have here dealt with, while leading to results which form a special case of the law of ancestral heredity, is not sufficiently elastic to cover the observed facts. The lesson

* 'Biometrika,' vol. 2, p. 220.

to be learnt from the present investigation is, however, that there is no essential repugnance between any of the main results of the biometric school and a theory of the pure gamete, but, on the contrary, it is perfectly possible to test such theories by biometric methods. We may fairly ask anyone who propounds in future a Mendelian or pure gamete formula as a general theory of heredity, to remember that it involves in itself definite laws regulating the reproduction of a population mating at random, and that it is incumbent on the propounder to test whether or not such laws are consistent with what we already know of the inheritance statistics of such populations. When we remember that deducing all the effects of such a formula within the whole field of inheritance will almost always form a very laborious piece of mathematical analysis, there seems a touch of scientific irresponsibility in propounding an immense variety of formulæ to suit one or other special case, and the modifying or withdrawing them when they are found to fail in another.

(10.) Proposition VIII.—*To find the Regression and Correlation of Brethren on the Theory developed in this Paper.*

We shall suppose the group of brethren to consist of 4χ members, or any pair of parents to have a family of 4χ.

Consider first parents of *one couplet only*, the offspring of the 16 possible pairs are given in the table below :—

		Father.			
		$a+a$.	$a+\mathrm{A}$.	$\mathrm{A}+a$.	$\mathrm{A}+\mathrm{A}$.
Mother	$a+a$	$4\chi u$	$2\chi(u+v)$	$2\chi(v+u)$	$4\chi v$
	$a+\mathrm{A}$	$2\chi(u+v)$	$\chi(u+2v+w)$	$\chi(u+2v+w)$	$2\chi(v+w)$
	$\mathrm{A}+a$	$2\chi(v+u)$	$\chi(u+2v+w)$	$\chi(u+2v+w)$	$2\chi(w+v)$
	$\mathrm{A}+\mathrm{A}$	$4\chi v$	$2\chi(v+w)$	$2\chi(w+v)$	$4\chi w$

Now let us take every pair of brothers out of each of these 16 families and form a correlation table of brothers. The following table gives the distribution of the various types :—

		First brother.		
		u.	v.	w.
Second brother	u	$4\chi(9\chi-4)$	$24\chi^2$	$4\chi^2$
	v	$24\chi^2$	$80\chi^2-32\chi$	$24\chi^2$
	w	$4\chi^2$	$24\chi^2$	$4\chi(9\chi-4)$

This gives a total of $256\chi^2 - 64\chi = 16 \times 4\chi\,(4\chi - 1)$ pairs of brothers, as it should, every brother in 16 families having $4\chi - 1$ brethren, and there being 4χ in each family. Clearly we can divide by 4χ, and we have the simplified form :—

	u_1.	v_1.	w_1.
u_2	$9\chi - 4$	6χ	χ
v_2	6χ	$20\chi - 8$	6χ
w_2	χ	6χ	$9\chi - 4$

where the subscripts 1 and 2 refer to the first and second brothers.

Let us simplify this by considering only the allogenic elements, η denoting either a heterogenic or a protogenic element. Then we have :—

	u_1.	η_1.
u_2	$9\chi - 4$	7χ
η_2	7χ	$41\chi - 12$

Thus the distribution of pairs of brothers in the case of the character being fixed by a single couplet is

$$(9\chi - 4)\,u_1u_2 + 7\chi u_1\eta_2 + 7\chi u_3\eta_1 + (41\chi - 12)\,\eta_1\eta_2.$$

This is to be read as follows: there are $9\chi - 4$ cases of both brothers being allogenic, to $41\chi - 12$ cases of neither brother allogenic, to $7\chi + 7\chi$ cases of one only of the two brothers being allogenic.

Now when we pass from a character fixed by one couplet to a character fixed by two, the above distribution can occur in either couplet, and every possible pair of brothers will be got by squaring the above expression. Proceeding in this way to n couplet characters, we have the following symbolic expression for the distribution of brothers

$$(4\chi)^n \times \{(9\chi - 4)\,u_1u_2 + 7\chi u_1\eta_2 + 7\chi u_2\eta_1 + (41\chi - 12)\,\eta_1\eta_2\}^n,$$

the omitted factor 4χ being restored.

This, I think, represents the distribution of a character measured by allogenic couplets in a population of pairs of brothers, *i.e.*, is the correlation table for brothers in the population—any term involving $u_1^p u_2^q$ representing when we put η_1 and η_2 equal to unity the number of pairs in which the first brother has p and the second q allogenic couplets in their constitutions.

Hence if we find the coefficient of u_1^p in terms of u_2, we shall have the array or

brothers to be found associated with a brother of p allogenic couplets. The above expression may be written

$$(4\chi)^n [\{(9\chi - 4)\, u_2 + 7\chi\eta_1\}\, u_1 + \{7\chi u_2 + (41\chi - 12)\, \eta_1\}\, \eta_2]^n.$$

The term involving u_1^p is

$$(4\chi)^n \{(9\chi - 4)\, u_2 + 7\chi\eta_1\}^p \{7\chi u_2 + (41\chi - 12)\, \eta_1\}^{n-p} \eta_2^{n-p} c_{n,\, p,\, o}.$$

Neglecting the constant factor, the distribution in u_2 is given by

$$\{(9\chi - 4)\, u_2 + 7\chi\eta_1\}^p \{7\chi u_2 + (41\chi - 12)\, \eta_1\}^{n-p}.$$

We require to find the mean of this array.

Put

$$\lambda = \frac{9\chi - 4}{16\chi - 4}, \qquad \mu = \frac{7\chi}{16\chi - 4}, \qquad \nu = \frac{4\,(5\chi - 3)}{7\chi}.$$

Then again, but for a factor independent of the power of u_2, the array may be read

$$(\lambda u_2 + \mu\eta_1)^p\, (\lambda u_2 + \mu\eta_1 + \nu\eta_1)^{n-p}.$$

The general term is therefore

$$(\lambda u_2 + \mu\eta_1)^{p+s}\, (\nu\eta_1)^{n-p-s} c_{n-p,\, s,\, o},$$

and we must sum from $s = o$, to $s = n - p$.

The general term has therefore its mean at the distance $1 + \mu\,(p + s)$ from $p + s + 1$ allogenic couplets, or its mean

$$= (1 - \mu)\,(p + s), \text{ allogenic couplets,}$$

and its total frequency $= \nu^{n-p-s} c_{n-p,\, s,\, o}$.

This gives a total frequency of the array proportional to $(1 + \nu)^{n-p}$.

Hence, taking moments, we have for the mean m of the array given by

$$m \times (1 + \nu)^{n-p} = \sum_{s=o}^{s=n-p} \{(1 - \mu)\,(p + s)\, \nu^{n-p-s} c_{n-p,\, s,\, o}\}$$

$$= (1 - \mu)\, p\, (1 + \nu)^{n-p} + (1 - \mu)\,(n - p)\,(1 + \nu)^{n-p-1}.$$

Therefore

$$m = \frac{1 - \mu}{1 + \nu}\, n + \frac{\nu\,(1 - \mu)}{1 + \nu}\, p.$$

Hence we see that

(i.) The regression between brothers is linear.

(ii.) The fraternal correlation which is equal to the regression $=$

$$\frac{(5\chi - 3)}{3\,(4\chi - 1)},$$

and is quite independent of the number of couplets. It is, however, a function of χ, the size of the family used in forming the table. We have the following values :—

Size of family.	Value of χ.	Value of fraternal correlation.
8	$\chi = 2$	·3333
12	3	·3636
16	4	·3778
20	5	·3860
24	6	·3913
32	8	·3978
40	10	·4017
∞	∞	·4067

The value of fraternal correlation thus varies with the size of the family dealt with from ·3 to ·4. Probably the more correct way of looking at any fraternal correlation table would be to suppose it a random sample of all the 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 brethren as they provide. In this case we ought to reach a fraternal correlation of ·4, precisely the value reached by the ancestral law when we take Francis Galton's original series of ancestral correlations.*

Thus we conclude that on the general theory of the pure gamete here dealt with, the fraternal correlation is slightly larger than the parental. This is in accordance with the general result of biometric investigations on populations. But the value, as in the case of the parental correlation, is very sensibly lower than the value—about ·5—found from recent investigations on man. It is further very inelastic even if we allow for some variation in the size of families dealt with. There can be little doubt that fraternal correlation varies from character to character and species to species in a manner sensibly beyond what can be accounted for by differences in the size of the family dealt with.†

Corollary.—We can exhibit the regression in the form :

$$\text{Mean of array} - \text{mean of general population} = \frac{\nu(1-\mu)}{1+\nu}\{\text{deviation of brother from mean of general population}\},$$

by observing that $1 - \mu = \frac{1}{4}(1 + \nu\mu)$, whence

$$m = \tfrac{1}{4}\frac{(1+\nu\mu)}{1+\nu}n + \frac{\nu(1-\mu)}{1+\nu}p,$$

* 'Roy. Soc. Proc.,' vol. 62, p. 410.

† There is sensible variation even for different characters, when we take the same series of pairs of brothers, and only one pair from each family.

or

$$m - \tfrac{1}{4}n = \frac{\nu(1-\mu)}{1+\nu}(p - \tfrac{1}{4}n).$$

(11.) PROPOSITION IX.—*To find the General Formula for Biparental Regression on the Theory of the Pure Gamete, and the Value to be given to the "Midparent."*

If we applied without further consideration the general formula for biparental regression to this case, we should have, if m_{pq} be the mean of the offspring due to fathers of p-allogenic couplets, mated with mothers of q-allogenic couplets,

$$m_{pq} = \tfrac{1}{4}n + \tfrac{1}{3}(p - \tfrac{1}{4}n) + \tfrac{1}{3}(q - \tfrac{1}{4}n).$$

This follows at once, since the mean of the general population $= \frac{1}{4}n$, the regression coefficient for either parent $= \frac{1}{3}$, and there is no assortative mating.

Hence we should have

$$m_{pq} = \tfrac{1}{12}n + \tfrac{1}{3}(p + q).$$

Now suppose both parents of pure allogenic race, then $p = q = n$, and all the offspring will be of pure allogenic race, or we must have $m_{pq} = n$.

But the above formula gives

$$m_{pq} = \tfrac{3}{4}n,$$

which is not correct.

In other words, while the above formula gives the best plane to fit the array of points determined by the parental constitutions, that array of points does not truly lie in a plane. Or, although the simple regressions are linear, the compound regression is not truly planar. We have therefore to find its true form, and measure the amount of deviation from the truth involved in using a biparental formula of the type indicated.

Given a character resulting from n couplets, we require $4^n \times 4^n$ individuals, 4^n male and 4^n female, to form the whole possible system of random matings. In such a population there would be

$$c_{n,p,o}3^{n-p} \text{ fathers of } p\text{-allogenic couplets,}$$

and

$$c_{n,q,o}3^{n-q} \text{ mothers of } q\text{-allogenic couplets.}$$

The chance therefore of a mating of a p-allogenic father and a q-allogenic mother is

$$c_{n,p,o}\, c_{n,q,o}\, 3^{2n-p-q}/4^{2n},$$

and when n be even moderately large, this gets very small if p and q at all nearly approach n. For example, if n were 5, and father and mother were both pure allogenes, the chance of such a pure allogenic mating would be only

$$1/1{,}048{,}576,$$

or in a population of a million would hardly occur once.

Still keeping $n = 5$, take $p = q = \frac{4}{5}n = 4$. Then we have the chance of such a mating

$$= 1/4215$$

still extremely improbable.

Thus when n is even moderately large, pure allogenic matings are so rare that they have vanishingly small influence in the population at large. Even if n were 2, the chance of a pure allogenic mating is only $\frac{1}{256}$. These points must be borne in mind in what follows.

Consider first a father and a mother of one couplet each, their zygotes are either u, v or w, involving a gametic constitution $a + a$, $a + \mathrm{A}$ or $\mathrm{A} + \mathrm{A}$. We have the following scheme :—

Zygote of father.	Zygote of mother.	Number of matings.	Offspring.
u	u	1	$4u$
u	v	2	$2(u+v)$
u	w	1	$4v$
v	u	2	$2(v+u)$
v	v	4	$u+2v+w$
v	w	2	$2(w+v)$
w	u	1	$2(v+v)$
w	v	2	$2(w+v)$
w	w	1	$4w$

Hence it there be

1 allogenic couplet in father and 1 in mother, offspring $= 4u$,

1 . . . . . 0 . . . . . $= 4(u + 2v)$,

0 . . . . . 1 . . . . . $= 4(2v + u)$,

0 . . . . . 0 . . . . . $= 4(u + 4v + 4w)$.

Let us write

$$16j_0 = 16\left(\tfrac{4}{4}u + v + w\right), \qquad 16j_1 = 16\left(\tfrac{1}{4}u + \tfrac{1}{2}v\right), \qquad 16j_2 = 16\tfrac{1}{4}u.$$

Then consider the relation

$$(u^0\epsilon^0 + u'\epsilon^1) \times (u'^0\eta^0 + u'^1\eta^1) = 4^2\left(j_0\epsilon^0\eta^0 + j_1(\epsilon^1\eta^0 + \epsilon^0\eta^1) + j_2\epsilon^1\eta^1\right),$$

where ϵ and η are mere symbols, and 0, 1, etc., denote their powers. u, u' refer respectively to father and mother, and their powers denote the number of allogenic couplets in the zygotes of father and mother. Then the above is a symbolical relation which gives, by equating any power or product of ϵ and η on either side, the offspring of a pair of parents of definite constitution.

Now suppose the parents not to consist of a single couplet, but of n couplets, then the total distribution of offspring that we have given above for any couplet may occur

in each couplet, and each such distribution must be combined with every other couplet distribution. We then reach, dropping unnecessary indices, the general symbolic relation,

$$\begin{aligned}&(u^0 + u'\epsilon + u^2\epsilon^2 + \ldots + u^n\epsilon^n)\\ &\times (u'^0 + u'^1\eta + u'^2\eta^2 + \ldots + u'^n\eta^n)\\ &= 4^n \times 4^n \times (j_0 + j_1(\epsilon + \eta) + j_2\epsilon\eta)^n.\end{aligned}$$

Thus the array of offspring due to parents of zygotes with p and q allogenic couplets respectively—*i.e.*, to $u^p \times u'^q$—is the coefficient of $\epsilon^p\eta^q$ on the right-hand side, or in the expansion of

$$4^n \times 4^n \times (j_0 + j_1(\epsilon + \eta) + j_2\epsilon\eta)^n.$$

This may be written

$$4^n \times 4^n \times (j_1 + j_2\eta)\,\epsilon + j_0 + j_1\eta)^n.$$

Thus the coefficient of ϵ^p is

$$4^n \times 4^n \times (j_1 + j_2\eta)^p (j_0 + j_1\eta)^{n-p} c_{n,\,p,\,o}$$

We require to pick the coefficient of η^q out of this in order to get the array of offspring due to fathers of p, and to mothers of q, allogenic couplets. But this is clearly

$$\begin{aligned}4^n \times 4^n \times c_{n,\,p,\,o}\{&j_2^q j_1^{p-q} j_0^{n-p} c_{p,\,q,\,o}\\ &+ j_2^{q-1} j_1^{p-q+2} j_0^{n-p-1}(n-p)\, c_{p,\,q-1,\,o}\\ &+ j_2^{q-r} j_1^{p-q+2r} j_0^{n-p-r} c_{n-p,\,r,\,o}\, c_{p,\,q-r,\,o}\\ &+ \ldots\},\end{aligned}$$

or, more briefly,

$$4^n \times 4^n \times c_{n,\,p,\,o}\left\{\sum_{r=o}^{r=q}(j_2^{q-r} j_1^{p-q+2r} j_0^{n-p-r} c_{n-p,\,r,\,o}\, c_{p,\,q-r,\,o}\right\}.$$

We shall first find the mean and frequency corresponding to the r^{th} term as given above of this series. What we have to deal with is

$$(\tfrac{1}{4}u)^{q-r}(\tfrac{1}{4}u + \tfrac{1}{2}v)^{p-q+2r}(\tfrac{1}{4}u + v + w)^{n-p-r}.$$

We may write this

$$\frac{3^{n-q+r} \times u^{q-r}}{4^n} \times (\tfrac{1}{3}u + \tfrac{2}{3}\chi)^{p-q+2r} \times (\tfrac{1}{3}u + \tfrac{2}{3}v + 2\chi')^{n-p-r},$$

where $\chi = v$ and $\chi' = \frac{1}{3}v + \frac{2}{3}w$, and both may be put unity when we are merely finding the distribution of allogenic couplets.

Now the general term in the above expression is

$$\frac{3^{n-q+r} \times u^{q-r}}{4^n} \times (\tfrac{1}{3}u + \tfrac{2}{3}\chi)^{p-q+2r+s}(2\chi')^{n-p-r-s} c_{n-p-r,\,s,\,o},$$

and s must be taken from o to $n-p-r$. The frequency of this term is

$$\frac{3^{n-q+r}}{4^n}\,2^{n-p-r-s}\,c_{n-p-r,\,s,\,o},$$

and its mean is $q-r+\frac{1}{3}(p-q+2r+s)$ allogenic couplets.

The total frequency of the r^{th} term is therefore

$$\frac{3^{n-q+r}}{4^n}(1+2)^{n-p-r}=\frac{3^{2n-p-q}}{4^n},$$

which of course must ultimately be multiplied by the factorials in r omitted above. If m_r be the mean number of allogenic couplets in the r^{th} term, we have

$$m_r\times\frac{3^{2n-p-q}}{4^n}=\sum_{s=o}^{s=n-p-r}\left(\frac{3^{n-q+r}}{4^n}\,2^{n-p-r-s}\,c_{n-p-r,\,s,\,o}\,\{q-r+\tfrac{1}{3}(p-q+2r+s)\}\right.$$
$$\left.=\{q-r+\tfrac{1}{3}(p-q+2r)\}\,\frac{3^{2n-p-q}}{4^n}+\tfrac{1}{3}(n-p-r)\,\frac{3^{2n-p-q-1}}{4^n}\right).$$

Thus:

$$m_r=q-r+\tfrac{1}{3}(p-q+2r)+\tfrac{1}{9}(n-p-r),$$
$$=\tfrac{1}{9}n+\tfrac{2}{3}q+\tfrac{2}{9}p-\tfrac{4}{9}r.$$

This is the mean of the r^{th} term, and its total frequency is

$$4^n\times 4^n\times\frac{3^{2n-p-q}}{4^n}\,c_{n,\,p,\,r}\times c_{p,\,q-r,\,o}.$$

Hence, if

$$\mathrm{F}=4^n\times 3^{2n-p-q}\sum_{r=o}^{r=q}c_{n,\,p,\,r}\times c_{p,\,q-r,\,o}$$
$$=4^n3^{2n-p-q}\times f,$$

we shall have

$$f\times m_{pq}=\sum_{r=o}^{r=q}\{(\tfrac{1}{9}n+\tfrac{2}{3}q+\tfrac{2}{9}p)-\tfrac{4}{9}r\}\,c_{n,\,p,\,r}\times c_{p,\,q-r,\,o},$$

where m_{pq} is the mean of the array of offspring due to fathers of p and mothers of q allogenic couplets.

The only difficulty here is summing the series

$$\sum_{r=o}^{r=q}\{r\,c_{n,\,p,\,r}\times c_{p,\,q-r,\,o}\}.$$

But this may be written

$$\sum_{r-1=o}^{r-1=q-1}(n\times c_{n-1,\,p,\,r-1}\times c_{p,\,q-1-r+1,\,o}),$$

or multipled by $4^{n-1}3^{2(n-1)-p-(q-1)}/n$ it represents the total offspring of fathers of p and

mothers of $q-1$ allogenic units in a population with $(n-1)$ couplets in their constitution

$$= 4^{n-1} \times c_{n-1,\,p,\,o} \times c_{n-1,\,q-1,\,o}\; 3^{2(n-1)-p-q-1}.$$

Hence

$$\sum_{r-1=o}^{r-1=q-1} \{n \times c_{n-1,\,p,\,r-1} \times c_{p,\,q-1-r+1,\,o}\} = n \times c_{n-1,\,p,\,o} \times c_{n-1,\,q-1,\,o}.$$

Now

$$4^n \times 3^{2n-p-q} \times f = 4^n \times c_{n,\,p,\,o} \times c_{n,\,q,\,o}\; 3^{2n-p-q}.$$

Thus

$$f = c_{n,\,p,\,o} \times c_{n,\,q,\,o},$$

or, if f' denote the above series, we have

$$f' = \frac{q\,(n-p)}{n}\, f.$$

This leads us to

$$m_{pq} = \tfrac{1}{9}n + \tfrac{2}{3}q + \tfrac{2}{9}p - \tfrac{4}{9}\,\frac{q\,(n-p)}{n}$$

$$= \tfrac{1}{9}n + \tfrac{2}{9}\,(p+q) + \tfrac{4}{9}\,\frac{pq}{n}$$

$$= \tfrac{1}{9}\,\frac{(n+2p)\,(n+2q)}{n}.$$

This is a most remarkable result, for it shows that the regression surface is not a plane but a hyperboloid. Let us measure all the quantities in deviations from the mean of the general population, *i.e.*, put $m_{pq} = m'_{pq} + \frac{1}{4}n$, $p = p' + \frac{1}{4}n$, $q = q' + \frac{1}{4}n$. We find

$$m'_{pq} = \tfrac{1}{3}\,(p'+q') + \tfrac{4}{9}\,\frac{p'q'}{n}.$$

This formula reconciles at once the Mendelian and Galtonian positions. When the number of couplets is large, parents having a number of allogenic couplets comparable with n are vanishingly small in number. The standard deviation σ of the population is $\sqrt{3n}/4$ (see p. 57).

Hence we may write

$$m'_{pq} = \tfrac{2}{3}\left(\frac{p'+q'}{2}\right) + \frac{p'}{3\sigma}\,\frac{q'}{\sqrt{3n}}.$$

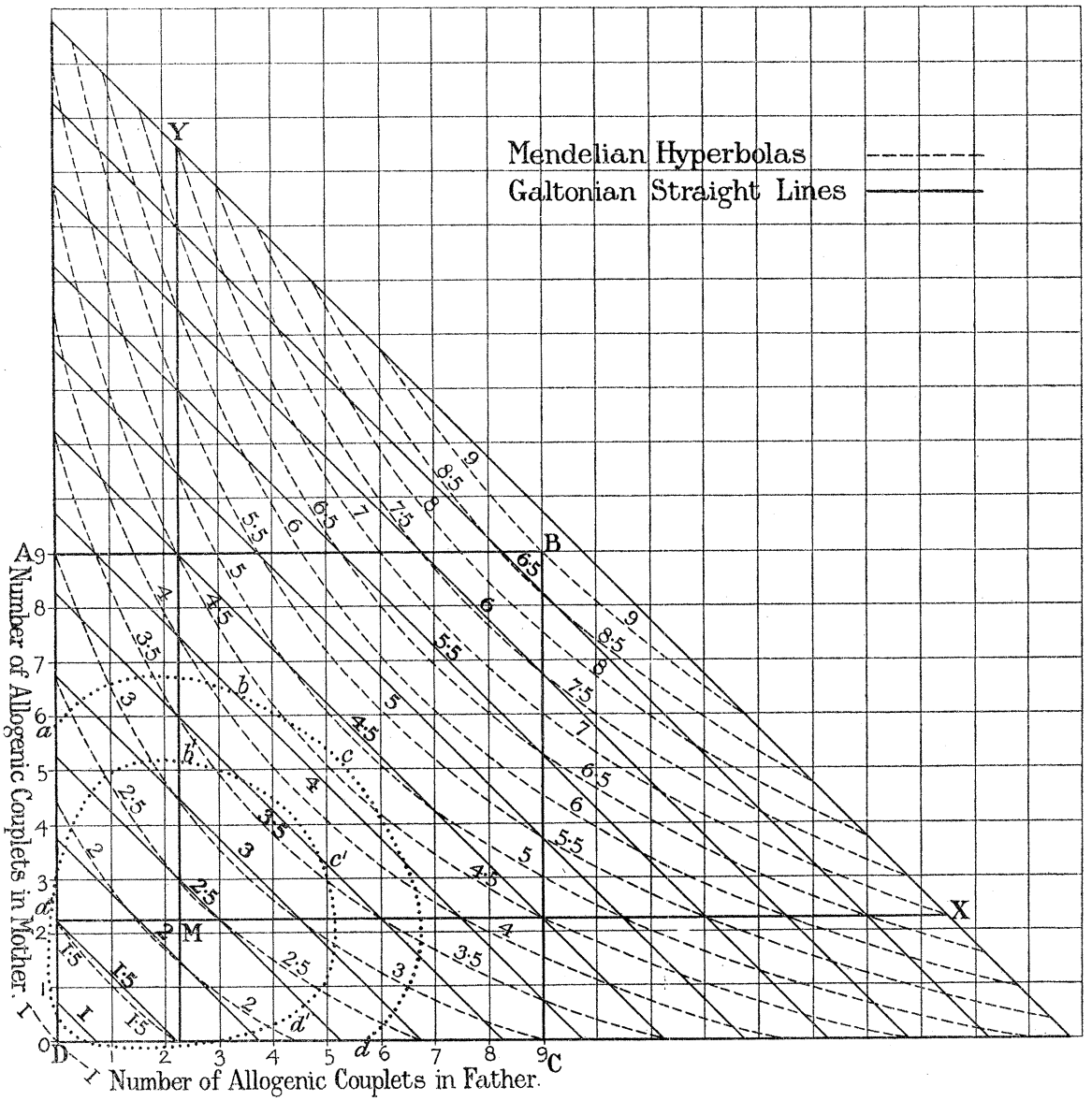
But p' can only once per thousand cases be as big as 3σ, and accordingly when n is large, both terms of the product will be small. In this case the surface becomes practically planar, and we have the Galtonian result of 1885,* that the offspring from the Galtonian "midparent" are one-third nearer the general mean of the population.

On the other hand, when n is small we see that for midparents not differing too widely from the population mean, Galtonian regression of the value $\frac{1}{3}$ holds, but that

* 'Natural Inheritance,' p. 97.

as we pass away from the mean of the population the regression gets less and less, becoming absolutely zero when we take two pure allogenic parents. Just as the regression is reduced when we emphasise the allogenic constitution of the parents, it is increased when we emphasise the non-allogenic elements.

To illustrate these points I have drawn a diagram for $n = 9$ of the contour lines of



Midparental contours.

the regression surface as a plane and a hyperboloid. The area ABCD contains all possible parents; 999 out of 1000 random matings fall in the loop *abcd*; 99 out of 100 random mating within the loop *a'b'c'd'*, which has been carried right round to mark that it excludes matings in which both parents have no allogenic couplets. The

means of the array of offspring due to any pair of parents are marked in slim figures along the Mendelian hyperbolic contours, and in heavy figures along the Galtonian straight lines. We see comparatively small differences as long as we deal with matings within the 99 per cent. loop, somewhat greater differences as we approach the 99·9 per cent. loop, and very marked differences as we go beyond the latter boundary towards pure allogenic parentage. A study of the diagram illustrates at once how the Mendelian theory exhibits for the bulk of the population Galtonian regression, such regression becoming, however, less and less as we proceed to individuals, the frequency of whose matings may be less than one in a million.

It will be seen again that in this proposition we have no fundamental antagonism between the Mendelian and biometric standpoints. We reach a single formula which approaches more and more closely to the biometric standpoint when we deal with characters depending on many allogenic couplets.* On the other hand, it gives the absence of regression which is obtained when pure allogenic parents are mated. We see, however, quite clearly that it is totally erroneous to argue from this single case against regression in general. Such regression actually exists on "Mendelian Principles" when any population breeding at random is taken, and involves in itself the whole conception of ancestral correlation and the influence of ancestry.

Of the formula for the midparent now reached, however, we can only say that on the basis of our experience in populations the factor $\frac{2}{3}$ seems too inelastic to work. Here again the data must be especially investigated from the standpoint of a midparent given by

$$\tfrac{1}{2}(p' + q') + \tfrac{2}{3}\frac{p'q'}{n}$$

before judgment can be final on this test.

Theoretically, by assuming the midparent to be

$$\tfrac{1}{2}(p' + q') + \chi p'q',$$

we should have a means of finding χ by averages, and therefore n, the number of couplets involved.

(12.) *General Conclusions.*

In this paper we have dealt with a general theory of the pure gamete—possibly not the widest that could be conceived—but sufficiently wide to indicate the real bearing of Mendelian formulæ when applied to a population mating at random. We see that under such circumstances:

(i.) The population which results from the offspring of hybrids remains stable; every variation which appears, appears with a certain definite and predicable

* For $n = 50$ or 100 the population within the 1 in 1000 line sensibly obeys ordinary linear midparental regression.

frequency; there is no room for the appearance of "mutations," although certain variations with very small frequency would be extremely rare in a limited population. A mutation—a variation not hitherto observed—would only appear in the offspring of the hybrids between two pure races; after this with random mating the mixed race would remain perfectly stable until disturbed by sexual or natural selection. These are the only mutations which arise on the generalised theory of the pure gamete, *i.e.*, *two* pure races form *one* mixed race, breeding true to itself; it is difficult under these circumstances to account for the origin of the two pure races by a mutation-theory of the differentiation of species!

(ii.) Between any two relations—if we measure the character by the number of allogenic or protogenic couplets in the zygote of the individual—we have a linear regression. The frequency distribution of any character is skew, approaching closely the normal distribution as the number of couplets which determines the constitution of the zygote is increased.

(iii.) The correlations between pairs of blood relations take definite numerical values absolutely independent of the number of couplets, and the same for all characters and races.

(iv.) The ancestral correlations form a geometrical series of common ratio one-half.

(v.) Fraternal correlation is fixed between narrow limits depending on the number of brothers per family dealt with, and is very slightly larger than parental correlation.

(vi.) The theory of the midparent for a considerable number of couplets approaches closely that originally given by Francis Galton, except for extreme values of the character, when the regression becomes rapidly smaller and ultimately vanishes.

We thus see that a generalised theory of the pure gamete would be of very great advantage if it could be accepted. It would lead to a system of inheritance in randomly mating populations with non-differential fertility, which in its broad features would be essentially the same as that which has been biometrically developed not from theoretical hypotheses, but from the statistical description of observed facts in populations.

Unfortunately, however, when we come to the actual numerical values for the coefficients of heredity deducible from such a theory of the pure gamete, they do not accord with observation. They diverge in two ways: First, they give a rigid value for these coefficients for all races and characters—a result not in reasonable accordance with observation. Secondly, they give values distinctly too small, as compared with the average values, or with the modal values of large series of population observations.

We thus reach the point we have so often had to insist upon: that the biometric or statistical theory of heredity does not involve a denial of any physiological theory of heredity, but it serves in itself to confirm or refute such a theory. Mendelian formulæ analytically developed for randomly mating populations are either consistent

or not with the biometric observations on such populations. If they are consistent, it shows their possibility, but does not prove their necessity. If they are not, it shows they are inadequate. The present investigation shows that in the theory of the pure gamete there is nothing in essential opposition to the broad features of linear regression, skew distribution, the geometric law of ancestral correlation, etc., of the biometric description of inheritance in populations. But it does show that the generalised theory here dealt with is not elastic enough to account for the numerical values of the constants of heredity hitherto observed.

It will be time enough to consider other more or less general Mendelian formulæ when there is far better evidence than exists at present that they cover a real range of observation, and have not been *solely* invented to describe isolated experiences, the numerical results of which are not in complete accordance with simple Mendelianism. Given such neo-Mendelian formulæ, there is a perfectly straightforward mathematical method of applying them to randomly mating populations, but that method is excessively laborious, and the biometrician may well hesitate to undertake the task of their investigation. A few minutes suffice to invent a Mendelian formula, but weeks of labour may be involved in testing whether it leads to legitimate results when applied to sexually crossing races. Let us therefore have a few simple general principles stated which embrace *all* the facts deducible from the hybridisation experiments of the Mendelians; these can form the basis of a new mathematical investigation, but it is idle to undertake such an investigation so long as Mendelian Principles remain in a state of flux.

Any combination of the theory of pure gametes here discussed with homogamy, or with fertility correlated with homogamy, or again with prepotency of individual or of type, would emphasise the correlations which we have found above to be too low; but such hypotheses would involve a fundamental alteration in the formula

$$(a + a')(\mathrm{A} + \mathrm{A}') = a\mathrm{A} + a\mathrm{A}' + a'\mathrm{A}' + \mathrm{A}\mathrm{A}'.$$

Such a formula would then give the *possibilities* of the cross, but the proportions of these possibilities actually occurring would be quite different.*

Such loading of the possibilities—not only of the individual couplet—but very probably of associated couplets in the constitution—might conceivably enable us to deduce better values for the ancestral and collateral correlations. But it would abolish not only the simplicity of the fundamental Mendelian formula, it would also involve lengthy preliminary studies on homogamy, fertility, and prepotency before any effective formula could be propounded.

* Toss two pennies, and the result of $4n$ tossings will closely approximate to the distribution $n(\mathrm{HH} + 2\mathrm{HT} + \mathrm{TT})$. Load one or both coins, and the possible variations will still be HH, HT or TT, but their proportions will be far from $n : 2n : n$.