

# MISCELLANEA.

## I. Inheritance in *Phaseolus vulgaris*.

Professor F. Johansen has just published a summary of his recent experiments on inheritance. His work *Ueber Erblichkeit in Populationen und in reinen Linien* (Fischer, Jena, 1903) is dedicated to Francis Galton, as the "Schöpfer der exakten Erblichkeitslehre"; it shows that the author has realised the importance of adequate statistical methods in any attempt to deal with the problem of inheritance, and we wish to express our gratitude to him for the courteous tone adopted in speaking of "Biometriker," and for the patient effort he has made to understand their work.

Professor Johansen's material is collected from three generations of beans: he has (1) a series of individual seeds, chosen out of a sample of about 16,000, which had been harvested and mixed together; these afford his evidence concerning the character of a grandmaternal generation; (2) each grandmaternal bean, when sown, yielded a crop of maternal beans, the mean character (weight or length-breadth index) of those borne by one plant being taken as the measure of the maternal character of the line of ancestry to which the plant belongs; (3) from the series of seeds borne by each mother plant, a certain number (from two to seven) were sown, the mean characters of the seeds borne by the resulting plants being taken as a measure of the characters studied in the filial generation of each line.

From these data three principal sets of tables are constructed; on p. 25, *Uebersichtstabelle I.* gives material for measuring the regression of individual filial beans on the mean weight of the maternal beans; a series of tables on pp. 21—24 gives the relation between the weights of the individual beans of the maternal plants, which were sown, and the mean weight of the seed produced by every resultant plant; finally on pp. 36—37 a table is given which Professor Johansen believes to show that the progeny of every seed, within a particular line of ancestry, exhibits a complete "Rückschlag" to the type of its line, the coefficients of correlation and regression between parent and offspring within the line being each = 0.

On the basis of these tables Professor Johansen attempts to explain the apparent discrepancy between Galton's law of regression and the results obtained by de Vries and others: but his view of the consequences supposed by "Biometriker" to follow from Galton's results shows that he has not fully realised what those consequences are.

It is fully realised by "Biometriker" that the general regression observed when we compare a filial generation with a parental generation is compounded of a series of sub-regressions, the members of each line of ancestry regressing to the "type" of their line; the effect of selecting definite ancestry for a small number of generations is also recognised; these points were fully dealt with in 1898, although few biologists seem to have realised the fact; it was then said:

"We now see that with the law of ancestral heredity.....a race with six generations of "selection will breed within 1·2 per cent. of truth ever afterwards,"\* or in other words the

\* K. Pearson: "On the Law of Ancestral Heredity," *Roy. Soc. Proc.* Vol. 62, 1898. Cf. pp. 397—402, "On the Variability and Stability of Selected Stock."

fixity of a line of ancestry is asserted when the ancestral purity is far less than that involved in Professor Johannsen's "reine Linie," which he defines as consisting of "Individuen, welche "von einem einzelnen selbst befruchtenden Individuum abstammen" (p. 9).

Again, "with a view to reducing the *absolute* variability of a species it is idle to select beyond "grandparents, and hardly profitable to select beyond parents. The ratio of the variability of "pedigree stock to the general population decreases 10 per cent. on the selection of parents, and "only 11 per cent. on the additional selection of grandparents. Beyond this no sensible change "is made."\*

In these two passages the fixity of the type and the high variability of individuals about the type are asserted as absolutely as they are asserted by de Vries in the passage of the *Mutations-theorie* quoted by Professor Johannsen. With our present knowledge of the coefficients of inheritance in man, horse, and dog, it would seem that from 2 to 4 generations of selection suffice to form a line varying greatly about its type, yet remaining true to that type. When we are told that a bean breeds true to its line we are told something which has been shown to be a necessary consequence of the law of ancestral heredity; if it were not true, the whole law would be upset.

The difference between the view put forward by Professor Johannsen, and that expressed in 1898 in the paper "On the Law of Ancestral Inheritance" is therefore not a difference which concerns the focus of regression in the offspring of selected ancestors; it is simply a difference as to the relation between successive generations of individuals within the line of ancestry. Professor Johannsen believes the tables on pp. 21-24 of his work to show that "*die persönliche Beschaffenheit der Eltern, Grosseltern, oder irgend eines Ahnen hat—soweit meine Erfahrung reicht—keinen Einfluss auf den durchschnittlichen Charakter der Nachkommen.* Es ist aber "*der Typus der Linie, welche den durchschnittlichen Charakter der Individuen bestimmt.....*" (pp. 61-62). Within the same line of ancestry, whatever individuals of a generation be chosen as parents, the character of the resultant filial generation will be the same according to Professor Johannsen, or the coefficients of correlation and regression between parents and offspring, within the same line of ancestry, will each = 0.

The experimental results do not seem to us consistent with this view. If the offspring of every generation, within a given line of ancestry, breed true to the type of their line, subject to such seasonal and climatic influences as affect the whole generation of their year, then when the whole filial generation is compared with the whole parental generation, correlation between the two must necessarily be perfect, and the coefficient of regression must be simply the ratio between the standard deviations of the two generations. Professor Johannsen should, we think, first have shown that perfect correlation does in fact exist between parents and offspring in two successive generations of plants; and this he has failed to do; in the case of his maternal and filial generations he has, however, published data which enable us to determine the required correlation, and the table below gives the result. With such data, the only method available is the first method used for Shirley Poppiest. It consists in determining the correlation between *every individual bean* of the filial generation, and the *mean character of its parent*. The absolute value of the correlation so obtained will not be significant, but the coefficient of regression will closely represent the true parental correlation,—being a relation between mean filial character and mean parental character. Taking Professor Johannsen's *Übersichtstabelle* I. (p. 25), and the maternal means given on pp. 21-24, the following table has been constructed, giving a coefficient of correlation = 0.531 and of regression =  $0.591 \pm 0.125$ . This latter value represents the coefficient of correlation, so far as the data allow it to be determined, between filial and maternal plants; and considering the paucity of maternal plants (only 19) the result is not in bad accord with previous results for parental correlation†. The value  $0.591 \pm 0.125$  is not very divergent from 0.5, but it cannot be held to approximate to unity!

\* Pearson: *loc. cit.*

† *Biometrika*, Vol. II. Part I. p. 69.

‡ *Biometrika*, Vol. II. p. 379.

TABLE A.

Correlation between Individual Filial Beans and Mean Maternal Beans.

Filial Beans.

Weights in milligrams	Filial Beans																Totals
	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80	80-85	85-90	
<i>A &amp; D 600</i>	—	—	—	5	2	11	26	47	82	98	84	46	26	17	6	2	452
<i>C 570</i>	—	—	—	—	—	5	14	50	76	58	44	29	5	1	—	—	282
<i>B 520</i>	—	—	—	—	1	6	19	32	66	88	100	90	50	19	1	3	475
<i>E 512</i>	—	—	—	—	4	1	12	29	62	65	57	19	6	—	—	—	255
<i>K 510</i>	—	—	1	2	6	31	55	55	28	6	4	—	—	—	—	—	188
<i>Q 440</i>	—	—	1	2	7	16	44	93	80	52	10	—	—	—	—	—	305
<i>R 410</i>	—	—	—	2	3	12	17	27	19	3	—	—	—	—	—	—	83
<i>S 405</i>	—	—	1	2	3	8	27	47	37	30	4	—	—	—	—	—	159
<i>G &amp; J 400</i>	—	1	5	23	66	155	283	353	241	97	15	—	1	5	—	—	1245
<i>F &amp; T 395</i>	—	—	—	2	9	27	66	111	85	58	22	1	1	—	—	—	382
<i>P 390</i>	—	—	—	3	1	18	35	27	13	3	4	2	—	—	—	—	106
<i>H 380</i>	—	—	1	6	20	60	106	114	75	33	3	—	—	—	—	—	418
<i>L 360</i>	—	—	1	5	15	37	88	76	33	13	4	1	—	—	—	—	273
<i>M 340</i>	—	—	4	9	26	56	82	76	32	9	1	—	—	—	—	—	295
<i>N 312</i>	1	3	11	22	29	72	120	69	23	5	2	—	—	—	—	—	357
<i>O 310</i>	4	4	5	19	69	69	44	5	—	—	—	—	—	—	—	—	219
Totals	5	8	30	107	263	608	1068	1278	977	622	306	135	52	24	9	2	5494

The letters *A* to *T* are used by Professor Johannsen to denote his nineteen "pure lines."

If Professor Johannsen's hypothesis were true, the only way in which he could account for the regression here observed,—a regression whose existence he himself admits,—would be by assuming that the characters, by which he has described the plants of his pure lines, are imperfectly correlated with the actual mean characters of his plants which he supposes to represent the types of his lines; but this assumption, while leaving his hypothesis logically unshaken, would destroy the whole value of his experiments as evidence in its favour, by destroying the value of the measure by which he has determined the characters studied.

Some of the difficulties we have felt in following Professor Johannsen are undoubtedly due to the imperfect way in which he has measured the characters of ancestral plants\*. The *grand-maternal plants*, for example, are determined each by the character of a single bean; the small value of such a determination may be judged from the variability among the beans of a single plant produced by the last generation; the mean number of such beans was 84.5, and the mean standard deviation of all arrays due each to a single plant was 75.37 milligrams, so that the mean character of a mother-plant, inferred from a single bean chosen at random among the offspring, would be equally likely to lie inside or outside the limits (true maternal mean + 50 mgrm.) and (true maternal mean - 50 mgrm.). There is thus, we venture to say, no strong probability that the numbers by which Professor Johannsen describes his grandparental beans represent the mean character of the seeds of the corresponding plants within  $\pm 100$  mgrm.

Again, the whole evidence, that the coefficient of filial regression within the line is zero, rests on the tables on pp. 21-24; but in these tables we are only told (1) the mean weight of seeds

\* A preliminary study of homotypy in the bean must of necessity precede any attempt to measure plant character by a single seed. Professor Johannsen would have to show that the homotypic correlation was perfect to justify his measure. This is very far from the fact not only in *Phaseolus vulgaris*, but in all beans hitherto examined from this standpoint.

produced by a mother-plant, representing the maternal character in a line; (2) the weights of individual seeds, taken from this mother-plant and sown; and (3) the mean weight of the seeds borne by each resultant plant. The regression coefficient, which Professor Johannsen regards (without any very adequate proof) as equal to 0, is between the deviation of the *individual beans* (2) and that of the mean character of the series of plants (3) resulting from them. We have therefore (1) the maternal generation defined by the mean character of all the beans borne by a plant; (2) a second generation, children of the maternal plant, each defined in the same way, by the mean character of the beans it bears; we have no third generation at all, and the regression which Professor Johannsen has observed seems to have little bearing on the question at issue, which could only be determined by growing a third generation from representative offspring of the filial plants, describing each plant of this generation as those of the two previous generations were described, in terms of the mean character of its seeds, and then determining the correlation between the characters so described in the two successive generations, the children and grandchildren of the single plant originally used to determine the line.

The question, which the tables given do to some extent answer, is the question what relation exists between the character of two seeds from the same mother-plant, and the character of the plants produced when those seeds are sown. Now it seems clear that if we take small beans out of a general harvest of seed, we shall be to some extent selecting the seeds of plants which bore on an average small beans; but what reason is there for supposing that the small and the large seeds from one and the same plant will lead to groups of plants bearing respectively small and large seeds? The hypothesis involved in this supposition seems somewhat analogous to the view that out of two eggs of a clutch, the smaller will produce a hen laying smaller eggs than that produced from the larger; this may well be quite fallacious, and it may yet be true that out of large masses of eggs, small eggs produce on the whole hens which lay small eggs. The absence of relationship of marked kind between the weight of seed sown, and mean weight of seed produced by the resulting plant, seems to have no bearing on the problem whether selection within the line can produce a change of character.

One further point we must notice; the *Uebersichtstabelle* 4 (pp. 36—37) has been treated in a quite illegitimate way, which would make the coefficients of correlation and regression = 0 between any two variables whatever.

If, as Professor Johannsen believes, the individual differences between members of any one generation were mere fluctuations, having no hereditary value, then a given generation ought as he says to be as well determined by selection of its grandparents as by selection of its parents. We cannot determine whether this is true of the average character of plants in Professor Johannsen's experiments, because the necessary data are wanting; but we can determine roughly the relation between three successive generations of *individual beans*. The material has been selected in such a way that the standard deviations of the successive generations have clearly quite artificial values, so that the correlations obtained are not very trustworthy; further, the exact weights of beans are not always given, so that we have been obliged to place a bean recorded as lying between say 400 and 450 mgrm. in the middle of its category. With these qualifications, we find

Correlation of Mother Bean and Offspring Bean  $r_{01} = 0.3481 \pm 0.0080$ ,

Correlation of Grandmaternal Bean and Offspring Bean  $r_{02} = 0.2428 \pm 0.0086$ .

If we wish to predict the weight of a given bean of the filial generation from the known weight of its maternal bean, we must form a regression equation, which becomes

Probable weight of Offspring bean =  $538.31 + 0.2691 \times \text{weight of Maternal Bean} \dots\dots(i)$ .

If we wish to predict the weight of a filial bean from knowledge of the grandmaternal bean, we obtain the regression equation

Probable weight of Offspring bean =  $417.41 + 0.1074 \times \text{weight of Grandmaternal Bean} \dots(ii)$ .

This shows us at once that from Professor Johannsen's own data the maternal bean is more than twice as influential as the grandmaternal bean in settling the weight of the filial generation.

If we correlate the selected maternal and grandmaternal beans, we find the correlation,  $r_{12} = 0.2532$ ; and from this and the preceding correlations we find the double regression formula

Probable weight of Offspring bean

$$= 330.71 + 0.2373 \times \text{weight of Maternal Bean} + 0.0731 \times \text{weight of Grandmaternal Bean} \dots (iii),$$

showing again the predominant influence of the maternal bean.

If we were to calculate the mean weights of each array of offspring means from (i), (ii) and (iii), we should expect the mean errors of the results to be in the ratio of

$$\sqrt{1 - r_{01}^2}, \sqrt{1 - r_{02}^2} \text{ and } \sqrt{\frac{1 - r_{01}^2 - r_{02}^2 - r_{12}^2 + 2r_{01}r_{02}r_{12}}{1 - r_{12}^2}},$$

or in this case as

$$1.014 : 1.050 : 1.$$

We have applied (i) (ii) and (iii) to the 65 arrays of offspring given by Professor Johannsen, and the mean errors are

$$44.3; 45.5, \text{ and } 42.8$$

or in the ratio of

$$1.035 : 1.060 : 1.$$

These numbers are, perhaps, as close as we could expect, and they show that we do in fact get better results from a knowledge of maternal bean than from knowledge of grandmaternal bean, and better results from a knowledge of both together than from a knowledge of either alone.

We hold therefore that Professor Johannsen's results prove :

(1) That there is a regression from parent to offspring, leading to the inference that parental correlation has for *Phaseolus vulgaris* a value closely identical with that found for other animals and plants, when we compare mean parental and mean filial characters;

(2) That when we compare the characters of individual seeds in successive generations the correlation between a seed and its parental seed is so much greater than that between a seed and its grandparental seed (both belonging to the same pure ancestral line) as to give strong evidence that characters arising in one generation within the line are inherited, and do therefore afford a basis on which selection may act.

W. F. R. W. AND K. P.

## II. Addendum to "Graduation and Analysis of a Sickness Table"

(*Biometrika*, Vol. II. p. 260). By W. PALIN ELDERTON.

On p. 261 it is stated that Gompertz' hypothesis may be viewed as the quotient of two normal curves, and it will be interesting to see how Makeham's useful modification of Gompertz' theory may be stated from the same point of view. Makeham's hypothesis is that the force of mortality may be represented as  $A + Bc^x$ , which, from our point of view, means that if we take a normal curve as the exposed we get  $(A + Bc^x) y_0 e^{-x^2/2\sigma^2}$  for the deaths, and this can be thrown into the form of two normal curves of the same standard deviation, viz.,

$$\frac{N_1}{\sigma \sqrt{2\pi}} e^{-x^2/2\sigma^2} + \frac{N_2}{\sigma \sqrt{2\pi}} e^{-(x-h)^2/2\sigma^2}.$$