THE DISTRIBUTION OF PURE LINE MEANS

DR. J. ARTHUR HARRIS

Carnegie Institution of Washington

Several times recently we have been told that the means of a character in a series of pure lines form a "Quetelet's Curve."[1] Some of those responsible for this assertion seem to attribute a particular virtue to "Quetelet's Law," and to feel that the statement that the means of a series of pure lines form a chance curve furnishes uncontrovertible evidence for the genotype theory of heredity. The questions which interest the biologist are, first, whether the statement is true in the sense that it is made on a sufficient body of actual observations, and second, what is the general biological significance to be attached to it, if true.

But among these biologists the interpretation of the facts has apparently preceded the demonstration of the existence of the facts themselves. Now while it is not at all unlikely that the means of genotypes—if such entities in Johannsen's sense of the term do exist in nature—form a chance curve, it by no means follows that conversely a series of averages which can be arranged in a symmetrical variation polygon proves or even suggests the existence of differentiated pure lines or biotypes. Yet just such differences in means are being accepted and cited without criticism as valid evidence in support of Johannsen's sweeping generalizations.

A case in point is a paper by Roemer[2] on pure lines in peas. It is with regret that one criticizes Roemer's

[1] Compare, for example, in this connection: Nilsson-Ehle, *Bot. Not.*, 1907, pp. 113–140; Lang, *Zeitschr. f. Ind. Abst.- u. Vererbungsl.*, Vol. 4, pp. 15–16, 1910; Spillman, Amer. Nat., Vol. 44, p. 761, 1910; Pearl, Amer. Nat., Vol. 45, p. 423, 1911.

[2] Roemer, T., "Variabilitätsstudien," *Arch. f. Rassen- u. Gesellsch.-Biologie,* Vol. 7, pp. 397–469, 1910.

paper. It is an exceedingly laborious *Arbeit* and apparently done with scrupulous care. One who himself has experienced the labor of calculating a few tables of constants has sympathy for a worker who has industriously filled pages with them. But the tenability of the genotype theory is one of the most pressing of current evolutionary problems, and all available evidence must be scrutinized. Roemer's data are chosen for two very excellent reasons, the first of which is that of all of the men who have discussed the disposition of the means of pure lines in a "Quetelet's Curve," he is, so far as I am aware, the only one who has put on record sufficient data for a critical test of his conclusions. If without overtrying the case, as the lawyers have it, we can give the second reason, it is that Roemer's data and conclusions have been accepted as perfectly valid by genotype specialists. One of them, for example, says:

> The work is essentially a confirmation, with another plant, of Johannsen's epoch-making investigations on beans, though it lacks any extensive studies on the effect of selection within the pure line. The essential objective point of Roemer's research is rather to determine the biometric characteristics of pure lines as such in relation to the general population. Among the more important general results are the following:
>
> 1. The different biotypes in a population arrange themselves in frequency distributions in accord with Quetelet's Law.
>
> 2. No relation was found to exist between the variability of the biotypes (*i. e.*, variation within the general population) and variation within the pure lines.

Our problem is twofold. First, we have to determine whether Roemer is really justified in regarding his lines as differentiated. Second, we have to inquire concerning the critical value of his data as evidence in support of the genotype theory of heredity. Incidently we shall make the first of these problems serve as an illustration of the use of a coefficient of individual prepotency recently proposed in these pages.[3]

[3] Harris, J. Arthur, "A Coefficient of Individual Prepotency for Students of Heredity," AMER. NAT., Vol. 45, pp. 471–478. 1911.

II. The Problem of Differentiation in Roemer's Pure Lines

The method of Roemer's study was very simple. In 1908 a population of pea plants was grown from a sample of ordinary seed. In 1909 the offspring of each of a number of these plants was studied separately, and the means of several characters calculated. By a comparison of selected pairs of these means Roemer concludes that the several lines differ from each other, and by a seriation of all the line means he obtains the Quetelet's curve.

Such evidence as this can not be accepted. Every mean calculated on a sample of individuals is more or less untrustworthy as a measure of the character in individuals in general, because of the errors of random sampling, and in attaching significance to a series of averages this fact must be fully taken into account. It can not adequately be allowed for by a comparison of selected cases with their probable errors.

First Test. A Comparison of the Variability within the "Pure Line" with that of the "Population"

One of the tests of the presence of differentiated "biotypes," "genotypes" or "pure lines" within a "population" is the comparison of the intra-line with the population variability. If both be the same there is no justification in the assumption that the population is composed of a number of differentiated pure lines.[4] If the variability of the population is greater than that of the individual lines it may (or may not) comprise a series of "genotypes."

The reason for this is obvious. The standard deviation within the pure line, σ, describes only the differences occurring among the individuals of the group, while Σ, the standard deviation of the group, includes also the amounts by which the several lines are differentiated.

[4] This is, of course, under condition that the individuals of the several pure lines are not reared under conditions which tend to increase artificially their variability beyond that of the population.

Roemer does not give us the population standard deviations for the several characters in 1909 but only the averages, $m_1, m_2, m_3, \cdots, m_s$, and the standard deviations $\sigma_1, \sigma_2, \sigma_3, \cdots, \sigma_s$. We may approximate the desired constants very closely indeed[5] by the following method.

Let there be s samples or pure lines of $n_1, n_2, n_3, \cdots, n_s$ individuals each, with means $m_1, m_2, m_3, \cdots, m_s$, and standard deviations $\sigma_1, \sigma_2, \sigma_3, \cdots, \sigma_s$. These form the population $S(n) = N$, for which the physical constants Σ and M are desired.

The mean is clearly $M = S(nm)/S(n)$.

In calculating the S.D. we may take the first two rough moments, ν_1', ν_2', about any point we please and adjust by the familiar formula $\sigma^2 = \mu_2 = \nu_2' - \nu_1'^2$. If the moments be taken about 0[6] $\nu_1' = M$, and it is at once clear that for the population

$$\Sigma = \sqrt{\frac{S[n(m^2 + \sigma^2)]}{S(n)} - \left(\frac{S(nm)}{S(n)}\right)^2},$$

when S indicates a summation for all groups or lines.[7]

The population constants have been calculated by these formulæ for all the characters of Roemer's two large series. He has given population constants, M and Σ, for the 1908 series, the parents of the 1909 plants.

The two are conveniently laid side by side for comparison in Table I.[8] The data in hand hardly seem to justify detailed comparison with reference to probable

[5] There is no approximation in the formula. The accuracy in practise depends solely upon the trustworthiness of the original m's and σ's, and upon the number of decimal places retained in the arithmetical routine.

[6] For several advantages in doing this see AMER. NAT., Vol. 44, pp. 693–699. 1910.

[7] The application of the formula to Roemer's data is of course exceedingly laborious, involving as it does the determination and summation by pairs of 3,108 squares, and the summation of the products of their totals by the frequencies upon which they are based. The publication of a little tabulated data would have reduced many days' labor necessary for a critical test of his results to a few hours.

[8] The constants for 1908 are taken from Roemer's Table I. Those for 1909 are calculated by the formulæ given above.

errors. It will be noted at once that for all the characters the mean is higher in 1909[9]—indeed for some characters in the "Kapital Erbse" it is almost double that found in 1908! With one exception the standard

TABLE I

COMPARISON OF PHYSICAL CONSTANTS FOR 1908 AND 1909 POPULATIONS

Character and Comparison	Averages		Standard Deviations		Coefficient of Variation	
	Gelbe Viktoria-Erbse	Svalöfs Kapital-Erbse	Gelbe Viktoria-Erbse	Svalöfs Kapital-Erbse	Gelbe Viktoria-Erbse	Svalöfs Kapital-Erbse
Weight of Plant						
1908 Population	13.09	7.99	4.250	2.815	32.47	35.23
1909 "	20.82	20.39	6.568	7.127	31.54	34.95
1909/1908 Ratio	1.59	2.55	1.545	2.531	.97	.99
Length of Stem						
1908 Population	114.96	78.96	12.985	12.575	11.30	15.95
1909 "	136.81	158.42	16.331	20.163	11.93	12.73
1909/1908 Ratio	1.19	2.01	1.257	1.603	1.06	.80
Thickness of Stem						
1908 Population	24.03	20.50	2.766	2.081	11.50	10.15
1909 "	25.20	24.05	2.390	2.933	9.48	12.19
1909/1908 Ratio	1.05	1.17	.864	1.409	.82	1.20
Number of Pods						
1908 Population	4.59	5.62	1.364	1.805	20.72	32.12
1909 "	6.48	11.54	1.987	4.263	30.66	36.95
1909/1908 Ratio	1.41	2.05	1.456	2.361	1.48	1.15
Weight of Pods						
1908 Population	9.71	5.76	3.192	2.184	32.80	37.95
1909 "	11.28	10.26	4.290	4.407	38.03	42.98
1909/1908 Ratio	1.16	1.78	1.343	2.017	1.16	1.13
Number of Seeds						
1908 Population	19.64	24.02	6.267	8.418	31.91	35.05
1909 "	26.60	45.11	9.162	18.633	34.44	41.31
1909/1908 Ratio	1.35	1.88	1.461	2.213	1.08	1.18
Weight of Seeds						
1908 Population	7.63	4.56	2.569	1.833	33.67	40.21
1909 "	8.56	7.26	3.511	3.376	40.99	46.49
1909/1908 Ratio	1.12	1.59	1.366	1.841	1.22	1.16

deviations in 1909 are higher than those in 1908. Mean and standard deviation are generally closely correlated, and this doubtless accounts for the greater variability of the 1909 series. Possibly, however, the 1909 plants

[9] Roemer states that conditions for growth in 1909 were superior to those in 1908.

were grown under conditions more heterogeneous than those to which the 1908 plants were exposed. In the second case, the S.D. might be directly raised, *i. e.*, heterogeneity in the crop may be merely a reflection of heterogeneity in the substratum.

There is no way of determining whether Roemer's cultural conditions were more heterogeneous in 1909 than in 1908, but it must be noted that *in ten of the fourteen cases* the coefficient of variation is higher in 1909.

Two ratios are to be examined,

$$\frac{\text{Mean Pure Line Variability}}{\text{Parental Population Variability}},$$

$$\frac{\text{Mean Pure Line Variability}}{\text{General Population Variability}}.$$

Consider first the ratio of the mean pure line to the parental population variability. If the offspring of the individual parents are differentiated we should expect to find the mean variability of the pure lines less than that of the parent population, providing, of course, that innate tendencies are not obscured by environmental factors. Table II[10] gives the necessary data.

Now the remarkable thing about these standard deviations is that in the most cases the variability within the individual "pure lines" in 1909 is greater than that of a mixture of all the pure lines in 1908. The excess is very striking in several cases. Of the fourteen comparisons, thirteen show a higher variability within the pure line than in the population. For the "Viktoria"

[10] For 1908 the population Σ and C.V. are from Roemer's Table I. The 1909 population Σ and C.V. have been calculated by the formulæ given above. The mean pure line standard deviations have been taken from Roemer's Tables II–III. None of the constants have been rechecked, since the original data are not available. The mean value of 12.13 for thickness of stem in Table II is obviously a printer's slip for 2.13. The mean pure line coefficients of variation are from Roemer's Table X. These were not calculated by dividing the sum of the coefficients of variation of the individual pure lines by the number of lines, but by dividing the mean standard deviation of the pure lines by the mean average of the pure lines.

the average M.P.L./Parental ratio is 1.327 while for the "Kapital" it is 1.996. The "pure line" variability is thus from 30 to 100 per cent. in excess of that of the population.

TABLE II

COMPARISONS OF MEAN PURE LINE AND POPULATION VARIABILITY

Character and Comparison	1908–1909. Standard Deviations		1908–1909. Coefficient of Variation		1909–1909. Standard Deviations	
	Gelbe Viktoria-Erbse	Svalöfs Kapital-Erbse	Gelbe Viktoria-Erbse	Svalöfs Kapital-Erbse	Gelbe Viktoria-Erbse	Svalöfs Kapital-Erbse
Weight of Stem						
Population...	4.250	2.815	32.47	35.23	6.568	7.127
Mean Pure Line...	5.96	6.44	28.53	31.40	5.96	6.44
M.P.L./Pop. Ratio ...	1.402	2.288	.88	.89	.907	.904
Length of Stem						
Population...	12.985	12.575	11.30	15.95	16.331	20.163
Mean Pure Line...	13.80	15.98	10.05	10.05	13.80	15.98
M.P.L./Pop. Ratio...	1.063	1.271	.89	.63	.845	.792
Thickness of Stem						
Population...	2.766	2.081	11.50	10.15	2.390	2.933
Mean Pure Line...	2.13	2.04	8.45	8.46	2.13	2.04
M.P.L./Pop. Ratio ...	.770	.980	.73	.83	.891	.696
Number of Pods						
Population...	1.364	1.805	29.72	32.12	1.987	4.263
Mean Pure Line...	1.80	3.92	27.19	34.00	1.80	3.92
M.P.L./Pop. Ratio...	1.320	2.172	.91	1.06	.906	.920
Weight of Pods						
Population...	3.192	2.184	32.80	37.95	4.290	4.407
Mean Pure Line...	3.80	3.95	33.84	38.50	3.80	3.95
M.P.L./Pop. Ratio ...	1.190	1.809	1.03	1.01	.885	.896
Number of Seeds						
Population...	6.267	8.418	31.91	35.05	9.163	18.633
Mean Pure Line...	8.01	16.60	33.30	36.80	8.01	16.60
M.P.L./Pop. Ratio...	1.278	1.972	1.04	1.05	.874	.891
Weight of Seeds						
Population...	2.569	1.833	33.67	40.21	3.511	3.376
Mean Pure Line...	3.17	3.04	37.05	41.90	3.17	3.04
M.P.L./Pop. Ratio...	1.234	1.658	1.10	1.04	.903	.900

The explanation of this anomalous result is first to be sought in the higher means (with the associated higher variability) in the 1909 plants. Basing the comparison on the coefficients of variation in order to eliminate, in so far as possible, the influence of the means, we note that seven of the ratios are greater and seven are less than unity, while for all the mean is .936. On an average, therefore, the pure lines have 93.6 per cent. as much variability as the population.

The second comparison, that between the variability of the individual pure lines and the population which they form, can be made on the basis of the standard deviations alone since the means are the same. This comparison (the last two columns of Table II) shows that in both series and for every character the variability written, the line is less than that for the population. The lowest ratio is .70, the highest is .92 and the mean is .858. This test indicates that they are differentiated. This is, of course, the conclusion which Roemer drew from his selected individual comparisons.

Second Test. The Deviation of the Pure Line Means from the Population Mean

For characters measurable on a quantitative scale the test for the deviation of the offspring of an individual from its population is given by

$$(m - M) \pm .67449 \sqrt{\frac{\Sigma^2}{N} + \frac{\sigma^2}{n}\left(1 - \frac{2n}{N}\right) - \frac{n(M-m)^2}{N(N-n)}}.$$

Where m and M, σ and Σ, n and N are the means, standard deviations, and numbers of individuals for the family and the population, respectively.[11]

For reasons which will be apparent to the reader later, the data which are given us do not justify calculations to a high degree of refinement.[12] We therefore approximate in every point possible.

The expressions

$$1 - \frac{2n}{n}, \quad \frac{n(M-m)^2}{N(N-n)}$$

[11] AMER. NAT., Vol. 45, pp. 471–478. 1911.

[12] Furthermore, I have serious misgivings that Roemer's lines comprise so few individuals each that the coefficient suggested must be used with caution. It is not needful to consider the point in greater detail here.

may be disregarded, since with such relatively low values of n as those of Roemer's data the first is practically unity, while the second is generally insignificant and may always be neglected, having a maximum value of *circa* .010, and in the majority of the cases falling far towards zero. The values of Σ^2/N are given in Table III. In practically every case the inclusion of Σ^2/N in

TABLE III

VALUES OF Σ^2/N

Character	Gelbe Viktoria Erbse	Svalöfs Kapital Erbse
1. Weight of plant	.0233	.0292
2. Length of stem	.1442	.9204
3. Thickness of stem	.0031	.0049
4. Number of pods	.0021	.0105
5. Weight of pods	.0100	.0112
6. Number of seeds	.0454	.1997
7. Weight of seeds	.0067	.0066

formula would change the end result only insignificantly, and since we are working roughly it may be omitted throughout.

We have thus reduced the formula to the fundamental term $\sqrt{\sigma^2/n}$. Fortunately for us, Roemer has not used the correct formula for the probable error of the mean but has given Johannsen's "Mittlerer Fehler," $\sigma/\sqrt{n} = \sqrt{\sigma^2/n}$. We now simply determine $m - M$[13] and $(m - M)/(\sigma/\sqrt{n})$ for each of his lines. Table IV gives the ratios.

Apparently there can be no question concerning the reality of differentiation in Roemer's lines. If the differences $(m - M)$ were due purely to random sampling from a homogeneous population, we should expect the

[13] This second test was carried out before the first, and so before the actual population means and variabilities were available. The means used were Roemer's means of pure line averages, as given in his Table IX. These would be the same as the population means if all the lines had the same n. As a matter of fact, the agreement is very close in all. In two cases only did I modify Roemer's means—changing number of pods from 6.6 to 6.5 and weight of pods from 11.2 to 11.3, in the "Gelbe Viktoria Erbse."

TABLE IV

Ratio of Deviation to Probable Error	Character[14] in Gelbe Viktoria Erbse							Character in Svalöfs Kapital Erbse						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
— 15.5 to — 14.5	—	—	—	—	—	—	—	—	1	—	—	—	—	—
— 14.5 to — 13.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—
— 13.5 to — 12.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—
— 12.5 to — 11.5	—	—	—	—	—	—	—	—	—	—	—	—	—	1
— 11.5 to — 10.5	—	—	—	—	—	—	—	—	1	—	—	—	—	—
— 10.5 to — 9.5	—	—	—	—	—	—	—	—	1	—	—	—	—	—
— 9.5 to — 8.5	—	—	—	—	—	—	—	—	1	—	—	—	—	—
— 8.5 to — 7.5	—	—	—	—	—	—	—	—	—	—	1	—	—	—
— 7.5 to — 6.5	—	2	—	—	—	—	—	—	—	—	—	1	—	1
— 6.5 to — 5.5	—	—	—	1	1	—	—	—	1	—	1	—	1	—
— 5.5 to — 4.5	4	3	—	—	3	3	2	2	1	1	2	—	1	1
— 4.5 to — 3.5	1	5	3	4	1	4	4	5	6	2	5	5	2	3
— 3.5 to — 2.5	6	12	8	5	4	3	5	7	2	2	1	4	4	6
— 2.5 to — 1.5	13	12	18	13	15	18	14	8	8	15	7	10	11	13
— 1.5 to — 0.5	29	19	19	24	24	25	25	19	15	25	15	24	19	16
— 0.5 to + 0.5	24	13	29	31	28	22	24	26	18	30	36	17	22	19
+ 0.5 to + 1.5	25	21	13	23	22	24	23	26	25	12	21	28	27	30
+ 1.5 to + 2.5	8	14	10	9	9	13	12	9	13	13	12	14	15	12
+ 2.5 to + 3.5	5	6	9	2	5	1	4	3	7	3	4	2	4	4
+ 3.5 to + 4.5	1	2	4	2	1	2	1	—	3	2	2	1	—	—
+ 4.5 to + 5.5	—	3	1	1	—	—	—	1	1	1	—	1	1	1
+ 5.5 to + 6.5	—	2	1	—	—	1	—	1	1	1	—	—	—	—
+ 6.5 to + 7.5	—	—	—	—	—	—	—	—	1	—	—	—	—	—
+ 7.5 to + 8.5	—	1	—	—	1	—	1	—	1	—	—	—	—	—
+ 8.5 to + 9.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+ 9.5 to + 10.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+ 10.5 to + 11.5	—	—	—	—	—	—	—	—	—	—	—	—	—	—
+ 11.5 to + 12.5	—	—	—	—	1	—	—	—	—	—	—	—	—	—

ratios $(m - M)/\sigma_{(m-M)}$ to have a standard deviation of $1 \pm .67449/\sqrt{2s}$, where s is the number of lines involved. The constants are given in Table V.

TABLE V

Character	Gelbe Viktoria Erbse	Svalöfs Kapital Erbse
1. Weight of plant	1.738 ± .077	1.886 ± .087
2. Length of stem	2.620 ± .117	3.209 ± .148
3. Thickness of stem	2.008 ± .089	1.768 ± .082
4. Number of pods	1.717 ± .076	1.955 ± .090
5. Weight of pods	2.213 ± .098	1.839 ± .085
6. Number of seeds	1.857 ± .083	1.759 ± .081
7. Weight of seeds	1.895 ± .084	2.188 ± .101

Now we remember that Roemer's individual lines are represented by so few individuals that our formula may not give absolutely trustworthy results, and that in case

[14] Numbers of characters are the same as in Tables III and V.

it does not, the discrepancy is in favor of the pure linist. But comparing these values with 1.000 ± .044 for "Viktoria" and 1.000 ± .046 for "Kapital," I think we must admit that the evidence is strongly in favor of a differentiation in these lines of peas.

III. THE BEARING OF THESE DATA UPON THE GENOTYPE THEORY OF HEREDITY

To the conclusion that these pure lines of peas are differentiated among themselves and that speaking roughly their means "form a Quetelet's Curve," there can be no objection, although this conclusion is by no means justified by Roemer's own analysis of his data. The assumption that these facts lend any support whatever to the genotype theory[15] seems to me to rest not only upon the most slipshod reasoning, but upon a complete disregard of simple biological precautions.

The chief of the pertinent reasons follow.

First. *There is no Evidence of Line or Genotypic Constancy or Heredity.*

By definition the genotype is a rigid organic entity, distinguished by breeding true from generation to generation, with the exception of mutations which are completely inherited and fluctuations which are not inherited at all. The mean of the line remains the same from generation to generation (except for rises and falls due to periodic environmental changes); the variations around these means are absolutely incapable of inheritance.

Roemer and his supporters of course assume this to be true for the *Pisum* series dealt with, but no fragment of evidence is adduced to show either (*a*) that these means remain the same from generation to generation, or (*b*) that selection within the pure line is ineffective. The condition is even worse than this. Roemer measured the

[15] The cardinal points of the genotype theory have been tersely, and I believe fairly, summarized elsewhere in these pages. See AMER. NAT., Vol. 45, pp. 346–363, 1911.

characters of the parent plants which furnished the seed from which his various pure lines were grown, but he neither determines whether there is any parental correlation in the population nor gives us the data from which this might be determined. To be sure, one may determine from his data that the variability within the individual line is less than that of the population, but this might be as easily attributed to the much-scouted "Galton's Law" as to genotypic heredity; possibly it is to be referred to some factor quite outside inheritance.[16] In fine, it is impossible to determine anything at all about inheritance from Roemer's data.

Second. *The Possibility of Environmental Heterogeneity is not Adequately Taken into Account.*

One of the first precautions of the experimental breeder should be to make sure that the environmental conditions to which his materials are subjected are not so heterogeneous as to vitiate entirely any conclusions concerning innate factors. Two watch glasses of *Paramecia* may present highly different environmental conditions. Food and housing are important factors in the egg records of poultry. Peas differ from row to row, or within the row, because of slight differences in the substratum. But experimenters now-a-days are so obsessed with the idea of rigid "unit characters," "determiners" and "genes" that little attention is paid to environmental influences; they are so absorbed in "analyzing" the "germ plasm" that they forget to make proper allowance for factors which may so modify the soma—which is the only available index of the germ plasm of an individual—that it is misleading as a guide to hereditary tendencies.

From my own experience in experimental cultures it seems quite possible that the differences in Roemer's lines are due to lack of uniformity in the substratum. One is

[16] For instance, de Vries has several times suggested that the feeding of the parent plant may have an effect on the offspring. I personally do not believe that this could be a sensible source of differentiation in Roemer's peas, but it is one of the factors which should be taken into account by a critical student of the problem.

not much impressed with the consistency of his results. Not only are his means and standard deviations much higher in 1909 than in 1908, but the relative variabilities are also higher in ten out of fourteen cases. To be sure, he has used the conventional precautions. He describes his field as a "gleichmässiger humoser Lehmboden," and states that the fertilizer was mixed with soil before application and distributed as evenly as possible. But for an organism so responsive to environmental influences as the garden pea,[17] and in a problem of this delicacy, these precautions are not at all sufficient. It is quite clear[18] that the seeds from each parent were planted together in rows, and if the soil differed at all from one part of the field to another the tendency would be for this heterogeneity to induce a differentiation in the crop.

If it be urged that we do not *know* that the differentiation in Roemer's means are due to environmental heterogeneity, the reply is simple. It is the duty of those who claim genotypic rank for observed differences to prove that their results are not due merely to faulty experimental conditions.

Third. *One or Both of Roemer's "Populations" are in Reality "Pure Lines."*

By definition a pure line is the offspring of a single individual of a continually self-fertilizing or vegetatively propagating organism. They may or may not be differentiated. Genotypists assume that generally they are both differentiated and highly constant. Selection within the line is absolutely incapable of effecting any change.

Now the curious thing about Roemer's material is that at least one, and possibly both, of the two populations suitable for our examination are in reality pure lines. Whether the "Individualauslese" by which the "Gelbe Viktoria Erbse" was developed gave rise to a pure line depends largely upon the stringency with which it was

[17] Roemer observes that the pea is exceedingly susceptible to environmental influences.

[18] See Roemer, *l. c.*, pp. 404–405.

carried out. Concerning "Svalöfs Kapital Erbse" there can be no doubt. The original plant was selected at Svalöf in 1896. Whether a further selection within the offspring of this plant was made, Roemer was unable to determine, but of the ancestry of his seed he felt quite certain.

Truly this is an anomalous state of affairs! Analyzed by the best available statistical methods, Roemer's data certainly indicate that the lines studied are significantly differentiated. Pure-line specialists dispense with any statistical analysis at all and accept the data as "a confirmation . . . of Johannsen's epoch-making investigations on beans." Yet if the differentiation in these lines be due to anything other than faulty experimental conditions, the observations described destroy entirely the value of Johannsen's theory by showing that heritable variations may occur in great numbers in the pure line.

IV. Summary and Conclusions

1. The statement that the means of the pure lines of a population form a "Quetelet's Curve" is now being made by genotypists. If it is true that an apparently homogeneous population is composed of a large number of slightly differentiated genotypes, it seems *a priori* not unlikely that their means will be arranged according to "Quetelet's Law." The question which concerns the biologist is whether this is, as a matter of fact, the condition found in nature. The object of the present rather laborious study has been to test the validity of this assertion on the basis of available facts. Roemer's data for pure lines in peas are the only passably satisfactory published series available.

2. Such a problem has two phases. It is first necessary to determine by adequate statistical tests that the lines in question may be reasonably regarded as differentiated biologically—*i. e.*, that the differences between them cannot be explained as the errors of random sampling, such

as give one a low or a high hand at cards. It is then allowable to consider the biological interpretation of the differences.

3. Two tests for differentiation were applied: (*a*) the mean intra-line variability was compared with the population variability, and (*b*) the significance of the deviation of individual line means from the population mean was tested by a coefficient of individual prepotency recently suggested. Both of these tests indicate sensible and statistically significant differences between the lines. These differences may be said to be distributed according to "Quetelet's Law" as the term is loosely used by biologists.

4. This fact *per se* furnishes no evidence at all for the genotypic nature of the differences in Roemer's lines. Indeed, throughout Roemer's work there is no conclusive evidence of any kind concerning any problem of heredity. At least one (and possibly both) of his series of material is from his own explicit statements in reality a pure line. The difference observed within these lines and considered by him and other pure linists to be of genotypic value and a confirmation of Johannsen's results with beans are probably merely the result of faulty experimental conditions. If they are not, Roemer's evidence goes squarely against Johannsen's theory.

COLD SPRING HARBOR, L. I.,
July 19, 1911.