VOL. IV APRIL 1930 PARTS I & II

ON A NEW THEORY OF PROGRESSIVE EVOLUTION*

BY KARL PEARSON, F.R.S.

THE thesis that I wish to propose in this lecture is the simple one, namely that (i) every individual differs in several or many characters from the type of his race, (ii) these individual characteristics are not reproduced fully in his offspring, because the bulk of his ancestry must necessarily be of the racial type. This is the phenomenon of regression first noted by Francis Galton. But (iii), if inbreeding occurs, this regression not only quickly ceases, but also changes to a progression. In fact the intensity of heredity is such that with isolation and inbreeding any individual characteristic, or deviation from racial type, will be gradually emphasised and become a factor of progressive evolution. Thus all organisms if isolated are in a constant state of evolution, and it is only interbreeding and the selective action of environment which preserve a type. In other words, natural selection controls evolution, but the progressive urge is provided by heredity itself.

I am aware that it will be difficult to bring this view home to the biologist, because it calls upon him to give up much of the paraphernalia of his genetic thinking; my proposed view of evolution strikes at the root of such a conception as the "pure line," if by the purity of the line we signify indefinite secular stability.

It may appear to you late in the day to speak of a new theory of evolution, and yet some such theory seems to be needed. That great intellectual view of life, which, whatever forerunners he may have had, Darwin undoubtedly first forced men of science to recognise, does seem at the present time at a standstill. The intensive work in all branches of biological science which flowed from Darwin's conceptions has begun to halt, not so much from a disbelief in the main outlines of the evolutionary hypothesis, but from the many difficulties which arise when we endeavour to picture mentally the origin and interaction of the factors, which together constitute the process we term evolution by natural selection.

The three familiar factors which are contributors to the process of evolution are: (i) Variation, (ii) Heredity, (iii) Destructive Selection, and the survival of those individuals best adapted to their environment. At first sight it appears so straightforward: living forms vary, and variations are inherited, yet as a rule the number which exists of any organism remains very nearly constant. When it is not so, the conditions of life for all associated organisms are bound to change rapidly. There are six sparrows, a cock robin and his wife, two chaffinches, and swallows building three nests to a cottage and its garden, but although each pair multiplies itself two to four fold in the year, yet when spring returns there are six sparrows, not twenty-one, one cock and hen robin, not three pairs, two chaffinches, not six, and three swallows' nests, not seven. A very little arithmetic will convince anyone of the heavy mortality among birds, and it is hard to believe it is not a selective mortality. In closely inbred races of dogs some 30 to 40 % of the puppies may perish; the feebler are pushed away from the mother's teats by the stronger, or would probably under any circumstances be non-viable. If you have a garden it is easy to observe the same struggle for existence between plants. Sow poppy seed thickly and innumerable poppies spring up, but before they flower you

---

* Based on a lecture delivered on January 29, 1929, at the Galton Laboratory.

will find only a few have survived to maturity. Or again, plant violets and montbretia and observe what happens in two or three years' time. Either coarse grass has killed the violets and montbretia, or you will have killed them by endeavouring to get rid of the grass. Or let us even take man himself. On the average the woman A who dies at 70 leaves more children behind her than the woman B who dies at 60, and yet neither A nor B have had any children since they were 48, and those of A have been longer exposed to risk. The explanation is that longevity is inherited and the longer living mother has longer living and therefore more surviving children. Now the degree of inheritance of measurable characters in man has often been ascertained, and it does not vary largely from character to character, but longevity has an inheritance falling below the intensity with which stature and span are inherited; this arises from the fact that *all* deaths are not selective. From the reduced inheritance in the matter of longevity, we are able to measure the intensity of the selective death-rate in man, and we find that 50 to 80 % of deaths in the case of man are selective*; the lower rate depends on the difference of environment as between parent and offspring, and the higher rate on the greater likeness of environment in the case of brethren. It is possible to measure the intensity of natural selection in the case of man much more easily than in other organisms because the necessary data have been long recorded. It is only in quite recent days that data on length of life have been experimentally determined for other forms of life, and their mortality table computed. I do not think it needful to pursue this point further, because the evidence that the major part of the death-rate in man (notwithstanding his apparent control of his environment) is selective seems to me overwhelming, and if in man, it is doubtless still more active in lower forms of life. The difficulty lies not in the existence of selection, but in determining how variations rise to have selective importance. Shortly, the problem of evolution lies in the discovery of the urge towards progressive variation.

There are two points occasionally overlooked when we are dealing with heredity, variation and the selective death-rate. The first arises from the distinction between the germinal and somatic cells of the organism. The former cells contain all the latent elements, which the individual can contribute to the next generation and which control, as far as depends on him, the somatic bodily characters of his offspring. But as the germinal characters are latent they cannot be directly subject to the selective action of the environment. Nature works indirectly on the germinal cells of the future generation by selecting the somatic cells in the existing generation. To illustrate this point let us suppose for the nonce albinism in man to be a simple Mendelian recessive. About one individual in 20,000 is a patent albino in this country; about one person then in 283 will, on the hypothesis of random mating, have a latent factor for albinism, but his bodily characters are to all appearance normal. The stock of human albinos is chiefly maintained by the intermarriage of stocks the germ plasms of which carry latent factors for albinism; it is only to a most insignificant extent the product of albino marrying albino†.

The patent albino is physically and often mentally inferior to the normal individual. Not a few of these albinos are found in charitable institutions, some in idiot asylums. Clearly in a state of nature they would perish at a higher rate than the normal individual. Among the Kaffirs albino children were by tribal custom exposed in the bush by their mothers; it is still considered desirable to free the tribe from them, and they find their way into the missionary schools. Generally the death-rate for albinos is considerably higher than for the normal population. Nature can and does

---

* *Biometrika*, Vol. I, p. 75.

† Out of 22,829 albinos, 20,000 would be offspring of the parents with latent factors for albinism, 2828 offspring of an albino with a latent factor-bearing parent, and only *one* the product of two albinotic parents.

very forcibly select them. But the extermination of albinism by the destruction or segregation of patent albinos would be an endless task on the hypothesis of random mating. It is the latent germinal factor, not only in albinism but in many other pathological anomalies, which is the crux of the situation. Looking at the audience before me, I am unable to say whether any of you bear a latent factor for albinism—the odds are about 3 to 2 against it,—and Nature as a selective force is still more ignorant; she can only work on patent characters. There is only one way in which those latent characters can be rendered patent and exposed to the action of selection, it is by inbreeding. Once we recognise that natural selection can act only on somatic and not directly on germinal characters, we see how slowly Nature must proceed as compared with the artificial selector. The breeder has so much more insight than Nature for two reasons: first, he can test by actual inbreeding the germinal constitution of his animals, and secondly he will if he be wise study the ancestry of the animals with which he is dealing. Neither process would be of the least avail, if the average somatic characters of individuals were not correlated with their gametic characters. There is, indeed, a very high correlation, which, even on the simple hypothesis of Mendelian alternative dominant and recessive factors, amounts to ·8165. In reality, since all of us really judge the constitution of the germ plasm of the stirp ultimately by the patent characters of ancestry and descendants, we are basing our conclusions on the existence of this somato gametic correlation. Without its existence the action of natural selection would be futile; among existing individuals it might destroy the less and preserve the better fitted to the environment, but it would not be in any way controlling their further reproduction. The mechanism by which the germ plasms of the father and of the mother may combine to produce the germ plasm of their offspring is a highly important scientific problem, but it has little to do with the problem of evolution by natural selection. That problem turns on the extent to which selection of somatic characters in the parentage will modify the somatic characters in the offspring. The relation between somatic characters in parent and offspring we can measure by observation or experiment without holding any theory of gametic inheritance at all. But since this relationship is close we are certain not only of the near association of the germ plasms of parent and offspring but of the high value of what I have termed the somato-gametic correlation. Considering how much depends for evolution on this correlation, it is curious how little attempt has been made by theorists to find its value.

We may sum up my first point then as follows: In the case of man and in the case of artificially bred organisms, it is the pedigree from which we learn most as to the germinal constitution of a stirp. Indeed, when we say that in the latter case the constitution can also be studied by breeding, it really only signifies that we are extending the pedigree in a *directed* manner. In considering the fitness for survival of an individual, Nature can only take into her purview the somatic characters of that individual, her blind forces have no appreciation of the pedigree value of the germ plasm; hence the efficiency of natural selection depends essentially on the existence of the somato-gametic correlation.

I now turn to the second point I wish to emphasise. Practically all the organisms we know exist in immense numbers and have existed for scarcely measurable periods of time. Nature can therefore afford to work leisurely, and what she destroys and what she perpetuates are there in masses. In other words, the discussion of natural selection must turn on mass-changes. It is accordingly a statistical problem, or, if you prefer it, an actuarial problem, a discussion of selective death-rates, based not on latent germinal, but on patent bodily characters. And further, what we want to know is the influence of such selection on the bodily characters of the next generation; we want to know how in the mass the average or type of the next generation is changed by selection of the parent generation, or more generally by the selection of certain stirps for survival.

The problem is one which has interest for the breeder and the eugenist as well as for the student of evolution. It may be stated as follows: Given all the information we can collect as to the *bodily* characters of the parental stirps, what is the most probable type of their offspring? This is the problem which statistics has had to solve and has solved. We can take the bodily character of one or of both parents, and predict what will be (i) the average or type of the offspring, and (ii) the variance of the offspring round this type. We can do precisely the same with one or all four grandparents and predict the like constants for the offspring generation. Then we can turn to the great-grandparents and, if their bodily characters are known, take them also into consideration and predict with still less variance the character in the offspring. Finally every ancestor and every collateral relative of whom the bodily character is known can be made to contribute his quota of information towards the prediction of the character in the new generation. Observation shows that there is a measurable degree of resemblance between each individual ancestor or collateral and their descendant or relative. This measure is expressed by a correlation coefficient, in this case a "coefficient of hereditary likeness." If we use all the known relatives available in the pedigree we obtain a formula of multiple correlation, which enables us to predict the type and variation of a sibship of such known ancestors and relatives. When we deal only with relatives in the direct ascent, I have termed this multiple correlation formula the "Law of Ancestral Heredity." It is a purely statistical law; it is independent of any theory of germ-cell mechanism, and it applies solely to the averages of large numbers. It is thus peculiarly adapted to the study of those problems in which heredity plays its part in evolution.

It has been often suggested that the Ancestral Law is contradicted by the discoveries of Mendel and his fellows; it is needless to say that this cannot be the case, for the law does not depend on any mechanism of the germ plasma. On the Mendelian hypothesis we can build up the tables which determine the degree of resemblance between any ancestor and his descendant; doing so we find these correlations decrease in geometrical progression precisely as the Ancestral Law, based on observation, assumed them to do. On the basis of these correlations, we can form the multiple correlation formula*. How then has the supposed contradiction arisen?

In the first place because biologists have applied to isolated individuals a prediction which is only true of the type or the average of all individuals of the like ancestries. In the second place because geneticists have chosen to give the same class name to a number of individuals, whose characters have a range of intensities which are independently heritable. And thirdly because Mendelians have asserted that the parents, and the parents alone, determine the characters of the offspring. This assertion is perfectly true, if by parents we understand the gametic characters of the parents. To know these is to know the gametic characters of the offspring. But this does not involve any contradiction in the Ancestral Law, for if we use the ancestral correlations, the geometrical series involved by Mendelian theory for the gametic characters, the multiple prediction formula reduces to its first term, the parental term, or the prediction of the gametic constitution of the offspring is based solely on that of their parents. Here there is no opposition between the Ancestral Law and Mendelism.

But, as I have said, we are concerned in evolution with the somatic characters and we wish to know how they progress under the influence of continued selection. In this matter we the eugenists have greater efficiency than Nature, for while Nature does not know whether you or I carry a

---

* See *Phil. Trans.* Vol. 203, A, pp. 53–86, and *R.S. Proc.* Vol. LXXXI, B, pp. 219–224 and pp. 225–229.

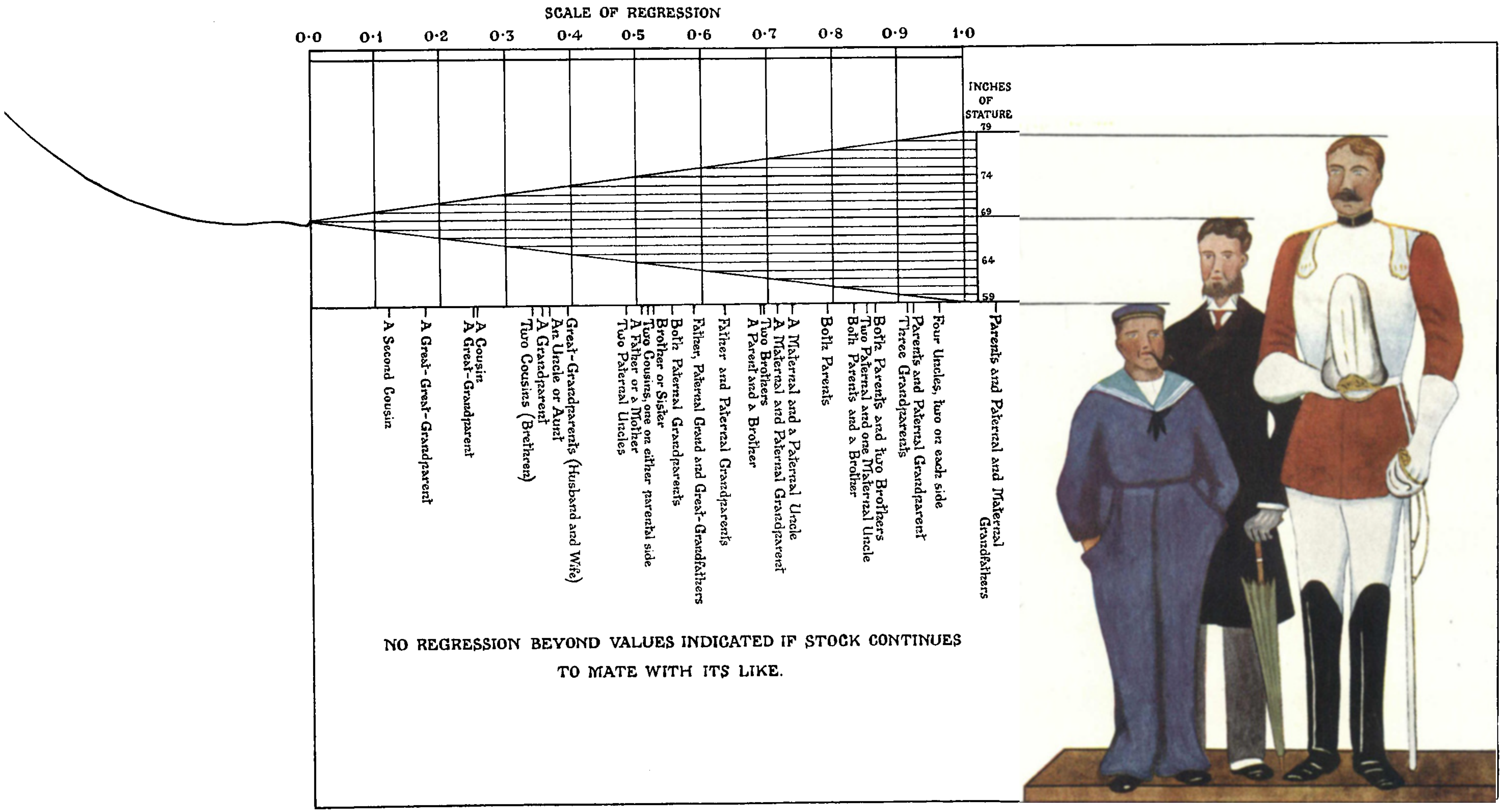
Take the red thread through any value on the scale of stature, say 74″, then the average stature of persons having all the kinsfolk described below of that stature would be obtained by drawing a vertical line through the mark indicated by the kinsfolk till it meets the red thread, and carrying through this meeting point a horizontal line back to the scale of stature, which provides the average desired. For example the average nephew of *two* uncles, one paternal and one maternal, each 6 ft. 2 in. would be 6 ft. 0·6 in., but the average nephew of *four* uncles, two on each side, each 6 ft. 2 in. will be 6 ft. 1·8 in.*. Again if both parents and paternal and maternal grandparents were of this stature, the grandsons would have progressed and be on the average 6 ft. 3·1 in. The statures are recorded for males, the corresponding female statures may be obtained by subtracting $\frac{8}{100}$ths from the male statures. In starting with females the male stature equivalent to that of the female must first be obtained by adding $\frac{2}{23}$rds of its value to the female stature. Thus a woman of 5 ft. 9 in. counts as a man of 6 ft. 3 in.

The reduction from my life-size diagram to the present small dimensions costs much in accuracy of reading, but serves to bring out the point that the regression ultimately changes to a progression.

---

* The regression coefficient used on the genometer was ·9614.

GENOMETER AFTER AN IDEA OF SIR FRANCIS GALTON



NO REGRESSION BEYOND VALUES INDICATED IF STOCK CONTINUES TO MATE WITH ITS LIKE.

latent factor for albinism, or insanity, a knowledge of our pedigrees may enable us to predict the *probability* of an albinotic child or of insane offspring.

I show you two red dogs, and you—like Nature—will be wholly uncertain, if you know nothing further, whether their offspring will be red or fawn, or black; or whether there will be some albinos among them. No Mendelian could tell you any more unless he had studied the pedigree upwards or downwards. Like the biometrician, he is reduced ultimately to a classification of his experience, and to the endeavour to get enough constants into a mathematical formula in order more or less adequately to describe that experience. Some day, perhaps, microscopic research may have advanced so far that from a close inspection of germ cells we may learn something of germinal potentialities. At present we are very far from such knowledge. In all cases we depend for our prediction on a knowledge of somatic characters in the ancestry, and further we cannot predict what will certainly be the nature of any individual, but only the type and variance of the whole array.

Now I do not propose even briefly to indicate the complicated mathematical process by which the statistician deduces his multiple correlation formula (see, however, the Appendix to this paper) but to place before you certain illustrations with some results which flow from them. I first draw your attention to a diagram (see Plate I) constructed many years ago from an idea which Francis Galton suggested to me. I based it on observations of over a thousand pairs of parents and their offspring and it is here worked out for stature; it might equally well have been based on span or cubit, which were also measured.

Now it is needful to remember an important point: When we take an exceptional relative and wish to obtain the *average* character of *all* men who have such a relative, we are supposing all his other relatives to run through the whole possible range of variation in this character; in other words, we give them an average or mediocre value. As we take more and more relatives of this exceptional degree of character, there is less and less of mediocrity in the remainder; thus the type of the array of men with these exceptional relatives rises higher and higher and the variance of the group becomes less and less. Actually the Ancestral Law allows every relative to produce his full effect on the prediction of the sibship by measuring his deviation from type by the variability of his sex in his own generation. For stature in man we can get rid of sex without this process, namely by reducing women to their male equivalents. Practically it is enough to add $\frac{1}{12}$ to the stature of a woman to convert it into the male equivalent and to take $\frac{1}{13}$ from a man's stature to reduce it to the female equivalent. To use the genometer the red thread is stretched across to the scale of stature at any definite value from mediocrity, and in order to predict the average stature of the sibship having the number of relatives indicated below with this deviation from type, we follow vertically upwards from the group indicated to the red thread and then move horizontally across to the scale of stature; where this horizontal line meets it we have the type of the sibship which has the given group of relatives of the given stature excess (or defect). Now if we examine the genometer we find that the selection of one exceptional parent leads to a very considerable drop back in the type of the offspring, that type has only about 51 % of the excess of the parent; but if we take both parents exceptional, about 79 %. This is the phenomenon, first noted by Galton, of *regression*. He thought there was always a continuous regression repeating itself in every generation. Unfortunately he was not mathematician enough to develop the multiple regression formula, or he would have discovered that there was no regression at all after this first stage, provided the offspring mate with each other or with their likes ("assortative mating"). It is random mating which would make the regression continuous back to the old population type. Such regression as occurs at first is due to the large number of relatives, who have not been selected. If we select both parents and one grandparent the offspring have 92 % of the selected character; and if we take both parents and the two grandparents on one side at the selected value the offspring, so far from showing any regression, have for type a 5 % advance on the selection value. To this point and its bearing on inbreeding I shall return shortly. Meanwhile I would impress upon you, that the amount of regression depends on the number of relations of selection value, and that beyond the first generation from selected ancestry, there will, if that generation inbreeds, mating with itself at random, be no further regression; the new population is stable.

The general principle thus stated holds equally good for Mendelian inheritance; if the offspring resulting from a hybridisation mate at random the population will remain stable, if there be no further selection. I proved this result for a multifactorial Mendelian system in 1903. It has been rediscovered by many Mendelians since and several times proclaimed as an important new result! To my mind this resulting stability is one of the difficulties which arise when we look for the origin of species in hybridisation. Bateson three years earlier, noting the great and striking variations which arise from hybridisation, wrote*:

> These new forms are like new species—some would even take rank as new genera. May not the natural species have arisen in like manner by hybridisation? The answer to this question is almost certainly *No*. And herein I believe most, if not all, professed botanists and zoologists will agree. To go into the matter fully here is impossible; but for many reasons, most of which have often been repeated, there is, I think, no good evidence for supposing that any natural species, whether of animal or plant, arose by direct hybridisation. Tempting as it may at one time have been to hope that we should thus get a short cut to the origin of species, few, I think, are now sanguine of such an issue.

Even to the end Bateson held that the Mendelian theory was not able to throw real light on the origin of species.

Now let us look at our genometer again. Having years ago worked it out for a number of cases, I saw how, by increasing the number of selected ancestors, I should ultimately reach offspring identical with the selected ancestors in character, and there would be no regression if the offspring interbred. The type would remain at the selection value. With this I was contented, and the matter remained on that footing for many years.

But parallel to this work on heredity in man, I have been carrying out for more than 25 years experimental breeding in dogs. I am not going to speak about that now, but simply to tell you of the origin of an idea. We started our dog experiments by crossing pure-bred albino Pekinese with pure-bred black Pomeranians. Neither of these varieties has very long legs, but the Pomeranian has a somewhat longer and far more slender thigh bone than the Pekinese. Owing to want of space and to want of money we have had to extend our experiments over a long course of years and this must mean, as we cannot keep many dogs at once, much inbreeding. Please remember, indeed, that this inbreeding very frequently does occur with any breeders who cannot afford large studs. Now I have been a good deal puzzled of recent years by a progressive tendency to lengthen the leg. Length of leg was not in my original reckoning at all! I had limited myself to coat colour, eye colour and the size and shape of head. But this very ungainly length of leg has been continually obtruding itself, although I have paid no regard to it in mating. At first I thought it possibly due to the form of exercise which seems peculiar to dogs confined in pens, namely the continual jumping up the walls dividing the pens to look over the top, but the *progressive* character of the long-leggedness seems to cut out this solution, unless we accept Lamarck's hypothesis, which I for one am loath to do. Well, here is the puzzle: close inbreeding, in fact isolation, for in the course of 25 years I have only once introduced a new dog, and yet progressive change.

---

* *William Bateson, Naturalist*, p. 163, Cambridge, 1928.

Unfortunately I did not measure leg length, it is not a convenient measurement on the living, and the cost of keeping all the skeletons of our 500 and more dogs was beyond my means. But the general idea was forced upon me that if we were given isolation and inbreeding, then a character, if even only slightly in excess of the type value, would continue to progress without special artificial selection, or any selective death-rate. The force which preserves the type in any species is summed up in the words *random mating*, or the whole deviations in excess and defect of type interbreed and cancel out, thus perpetuating mediocrity, and keeping the type stable*. Stability generally, as on the Mendelian hypothesis, depends on random mating. Given isolation and inbreeding, say from a single isolated pair, then if both members of the pair differ by excess or defect from type—and this by however small a quantity—their offspring by continual inbreeding will progress for this character, and therefore for all correlated characters. This is the idea that came to me from the long-leg progress of my dogs. Let us see how it would affect our ancestral prediction as exhibited in the genometer. Hitherto I had been contented by supposing that if we selected ancestry far enough back we should ultimately get offspring practically identical in character with the selected ancestry. It had not occurred to me that we might ultimately reach a value in excess of the selected value and constantly increasing as more and more of the ancestry exceeded the type value, and the selected individuals interbred.

Let us look into the reason why the child of an exceptional father regresses on that father, putting aside the mother who may not be as exceptional as the father and will not on the average be, if there be random mating. Twenty generations ago each one of us, but for cousin or other blood marriages, would have over a million ancestors of the same generation, and that is only 600 to 700 years ago. You will see that the ancestors of any one of you must 1000 years ago have been nearly synonymous with the total population of Great Britain and therefore their average was mediocre or your mid-ancestor of that generation was identical with the type of the inhabitants of this country or rather of a still wider range of mankind. I have endeavoured to follow back the ancestry of Charles Darwin and in those lines in which I can follow it, I find that a thousand years ago his known ancestors were scattered over the whole face of Europe, and stretched over Russia to Persia and through Constantinople to Asia Minor. No escape from this burden of ancestry is possible either in man or beast except by isolation and inbreeding.

Let us examine a little more closely the 10th generation of ancestry; we each had 1024 ancestors, between 300 and 400 years ago. Their average must be very nearly mediocrity, or type, for 1000 is a fair sized sample anthropometrically, as you would appreciate if you attempted to measure 30 or 40 characters in 1000 skulls! Indeed I may bring it still further home, for it would be difficult to find a man with 16 exceptional great-great-grandparents, and the average even of 16 individuals, if they are selected at random, will not be very different from mediocrity for any single character. Isolation of single pairs and inbreeding seem to me the only way in which differences far too small to have selective value can ultimately progress till they or the correlated characters have vital or lethal weight.

Now let us see how these ideas fit in with what we know of hereditary characters in man. I have spoken of the multiple correlation formula which gives the prediction on the *average* of the offspring for any or all ancestors. The only case we know of individuals having absolutely the same ancestry are siblings—whole brothers or sisters—offspring of the same pair of parents. They have the same set of ancestors right back to Adam, or to the first primate that could qualify as human. Now it is possible to deduce the multiple correlation coefficient on which the prediction formula

---

* "Panmixia" does not signify, as Weismann held, regression, but stability of type.

from all ancestors depends by a simple consideration of the variability of the population at large for any character with the corresponding variability inside the family. The ratio of these two variabilities for a considerable number of characters in man takes a value near to ·67, or it is practically impossible to reduce variability in man by any amount of selective breeding by more than about 33 %.

A knowledge of this percentage enables us to obtain the multiple correlation coefficient, which in man turns out to be about ·74, and from this we can deduce the formula which provides the type offspring of any system of ancestry. This formula is of the following form:

$$\text{Offspring type} = c\,(G_1 + \eta G_2 + \eta^2 G_3 + \ldots),$$

where c and η are constants that can be determined from observation of the fraternal and parental correlations and the G's are functions of the characters in the ancestors of each successive generation. The G's are in actual observation and in special hereditary theories related one to the other, the correlation between G_p and G_q $(p < q)$ being given by a geometrical series $\alpha\beta^{q-p}$.

In order that you may not think this a mere cobweb spun by the mathematician, I note the following values which flow from the Mendelian theory:

Table of α, β, c and η.

Simple Mendelian alternative factors	α	β	c	η
(a) Gametic constitution	1	$\frac{1}{\sqrt{2}}$ (= ·7071)	1	0
(b) Somatic constitution	$\frac{2}{3}$ (= ·6667)	$\frac{1}{\sqrt{2}}$ (= ·7071)	·5617	·4384
(c) Actual observations on Stature in Man				
(i) Female offspring	·7193	·999,024	·04787	·951,167
(ii) Male offspring	·7128	·999,759	·02407	·975,663

Now let us interpret these symbols:

(a) In this case $\eta = 0$, or we have the statement that the gametic constitution of the offspring is solely determined by that of their parents, a well-known principle of Mendelism, which in the early days of Mendelism was supposed by its adherents to render the study of distant ancestry idle!

(b) Here the formula is

$$\text{Offspring type} = {\cdot}5617\,G_1 + {\cdot}2463\,G_2 + {\cdot}1080\,G_3 + \ldots$$

or for prediction purposes the ancestral generations are in Mendelism itself of importance, if of diminishing importance, as we ascend in ancestry. If all the ancestral generations have the same constitution, i.e. $G_1 = G_2 = G_3 = \ldots = G$,

$$\begin{aligned}\text{Offspring type} &= G \times ({\cdot}5617 + {\cdot}2463 + \ldots)\\ &= G\,\frac{{\cdot}5617}{1 - {\cdot}4383} = G\,\frac{{\cdot}5617}{{\cdot}5617} = G,\end{aligned}$$

or, we reach a fundamental result of Mendelism that with random mating the race is stable as far as somatic characters are concerned. Thus, as applied to the simplest forms of Mendelian theory, the Ancestral Law neither contradicts Mendelism, nor is it self-contradictory.

Now what happens in the case of our own observations on stature—and it is the same for other characters in man? Why β within the limits of exact observation is *unity*. To have got so near to unity as ·999 is surely to have reached some very important natural principle, and it is so. It denotes that if we consider the offspring G_0, and the individual groups of ancestors of each generation, all these, like a group of brothers, are equally related to each other. If you are going to predict the type of offspring from any single generation of ancestors you may do it equally well from both parents, or from the four grandparents, or from the eight great-grandparents or the 16 great-great-grandparents and so on. There is no dying out of influence. A *single* great-grandparent has much less influence than a single parent, but eight great-grandparents have as much influence as the two parents in determining the type of the somatic character of the offspring.

Now the mathematics of the multiple prediction formula swing over the moment $\beta = 1$, and become very easy of treatment; η becomes also unity, and if we take the prediction from n generations we find for stature in man:

$$G_0 = \frac{\cdot 5567}{1 + (n-1)\cdot 5567}(G_1 + G_2 + \ldots) \text{ for male offspring}$$

$$= \frac{\cdot 5616}{1 + (n-1)\cdot 5616}(G_1 + G_2 + \ldots) \text{ for female offspring.}$$

I do not think there is any real difference between the two, but I have worked them out independently, for they thus confirm each other. The moment I reached these formulae, I saw that they fitted in with all that was being thrust on me from other sources of investigation.

If you start with a parentage, however little in excess of type, for one, or, for several characters, and inbreed, *the type, so far from being stable, will progressively alter, without any selection whatever.*

Approach the matter from this side, then we see that there is no stability, if there be inbreeding. Stability means practically random mating; with isolation and inbreeding or with assortative mating there is continuous progressive change. Some of you may at once say: You don't mean that if you took a pair of organisms with some slight excess from type and inbred from the pair that you would get a race with the character ever larger and larger? No! I do not think I should, because it has been tried. The fact is that all the skeletal parts, say of man, are highly correlated, and if you select one limb you modify all the remainder, and with the skeletal parts you modify all the soft parts, and you reach an organism incapable of surviving in its given environment. But I do mean that by inbreeding you can reach an increase or decrease of a given character, which will only stop short at lethal values. I believe indeed that at least half of the discredit cast on close inbreeding is due to the influence it has in continuously emphasising the deviations, whether in excess or defect from type.

Some writers on familial diseases have stressed what they term antedating, i.e. the appearance of the disease at an earlier age in the children and grandchildren of the affected. Most of the evidence provided by medical writers on the problem of antedating is statistically worthless, but if antedating exists, and I think it possibly may exist, it would be consistent with the progressive change to which I have referred.

Now let us look into this progressive change a little more closely. We will suppose that the parentage G_1 has a change of h* from type which may be as *small* as you please. This deviation h may be due to the father, the mother contributing nothing, or even a deviation below type; or it may be the other way round, or both may contribute to h. We will further suppose all the back ancestry to have been on the average mediocre or true to type. The offspring F_1 will have an average deviation $c_1 h$. Now to obtain their offspring F_2, supposing them to breed together, we have to

---

* h is measured in the variability σ of the corresponding generation as unit.

remember that the F_2 generation have not only parentage of excess c_1h, but grandparentage of excess h. We will call their excess c_2h. The third filial generation will have parentage c_2h, grandparentage c_1h and great-grandparentage h. Thus, as we advance to each successive generation, we have more and more selected generations of ancestry, and the problem before us is: Does this increasing selection of ancestry outweigh what Galton would have termed regression to type?

Let us suppose the parentage has a deviation from type $= h$.

Then the filial F_1, I find, has a deviation $\cdot 699h$. In other words, we have Galton's regression, but, knowing nothing about multiple regression, he supposed that the F_2 generation would have a deviation $\cdot 699 \times \cdot 699h$ and so on, thus ultimately regressing to the population mean.

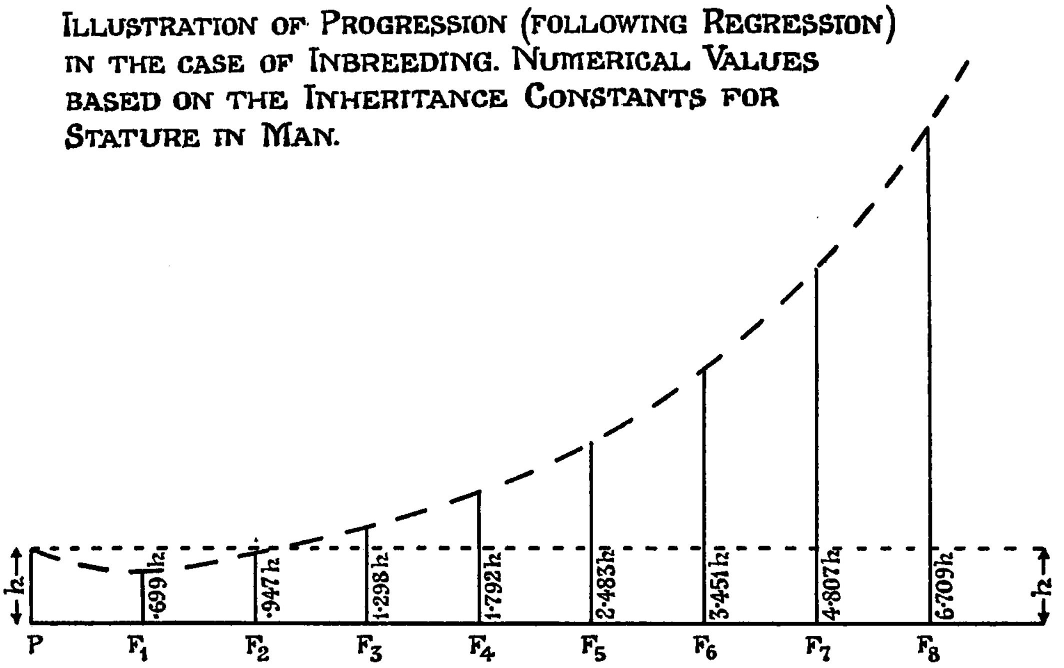


ILLUSTRATION OF PROGRESSION (FOLLOWING REGRESSION) IN THE CASE OF INBREEDING. NUMERICAL VALUES BASED ON THE INHERITANCE CONSTANTS FOR STATURE IN MAN.

$h\sigma$ is the parental deviation from type, F_1, F_2, F_3,.... represent the successive inbred filial generations, and the ordinates $x\sigma$ give the deviations from type of these generations. When the ordinates do not reach the deviation h of the parental generation, there is regression as in F_1 and F_2. When they exceed it as in F_3 and onward there is continuous progression as the result of inbreeding.

In doing this he forgot that the F_2 generation, if the F_1 inbred, would have a grandparentage of type h. Remembering this, we find for F_2 the deviation $\cdot 947h$.

Still regression, but not so great. Now the F_3 generation has parentage $\cdot 947h$, grandparentage $\cdot 699h$ and great-grandparentage h, hence we find for the deviation of F_3 $1\cdot 298h$.

We are now advancing beyond the original parentage, and, pursuing our formula further, we find

For F_4 the deviation $1\cdot 792h$
,, F_5 ,, $2\cdot 483h$
,, F_6 ,, $3\cdot 451h$
,, F_7 ,, $4\cdot 807h$
,, F_8 ,, $6\cdot 709h$

and so on. The series is represented graphically in the above diagram.

The point of these numbers is: that while we begin with the regression noted by Galton, yet since we have through the inbreeding more and more ancestors in excess of type, so the regression becomes ultimately a progression and, however small the original divergence from type, it is in the end magnified into significance. Thus we may say that it must become ultimately of selective value, whether by increasing fitness to the environment or by involving lethal changes in the organ itself or in correlated organs.

As I see the matter now, the individuals in any species are always tending to progressive change from type, and the population only remains stable owing to random mating. Divergence from type in every variety of characters, some with deviations above, others with deviations below mediocrity, constitutes what we term individuality, and accordingly with inbreeding a progressive evolution in the direction which emphasises individuality will always be started. It is true that assortative mating would achieve the like result, but much more slowly; it would, however, only be for a portion of the individuality, the sexually selected character and those organs which may be correlated with it. We must always bear in mind that in the case of man we have quite significant, if small, coefficients of assortative mating, for the fairly numerous range of characters so far investigated. The special advantage of inbreeding above assortative mating, even if we disregard its greater rapidity, is that it emphasises, not a selection of characters but the whole complex of factors constituting individuality.

It is true that thus far I have only indicated to you that the numerical values found for certain characters in man strongly suggest that inbreeding would be followed by progressive change. Let us accept for the time being this suggestion and see what difficulties such an hereditary urge would help us to solve. Many years ago a strong attack was made by Sir Edward Fry on Darwinism on the ground that the first appearance of a deviation from type must be too slight to be of selective value. He illustrated his argument by appeal to cases of mimicry, wherein it could not be supposed that the entire pattern would be developed as a single variation. It must, he held, have been developed gradually; but in the earliest stages, it would have, he said, no protective value, and accordingly the supposed gradual change by aid of continuous natural selection could not have taken place*. The Darwinians, among whom was Francis Galton, did not, I thought, come out of the controversy with flying colours. What at the time seemed to me to be needed, was an urge, once a deviation was started, for the progeny to continue with acceleration in the same direction. I believe this is possible, provided we inbreed. I hold that it is this inbreeding rather than directed selection—which of course may help matters—that has led to new varieties being produced even by breeders with relatively small stocks. I should look for varieties good or bad in domesticated animals in secluded mountain valleys served by a very limited number of home-bred stallions, bulls or rams.

---

* With progressive evolution the search for a definite use-value for *every* differentiation between local races becomes to a large extent idle. Many such searches have proved only the ingenuity of the solution-propounder. Of course certain characters can be directly correlated with habit or with environment, and are good evidence for the influence of natural selection. But it is, admitting progressive change under inbreeding or the isolation of small groups, no longer any argument against natural selection to say no purpose has or can be found for this or that difference between local races. From this standpoint a paper by Miss Jean M. Linsdale on "Varieties in the Fox Sparrow (*Passerella iliaca*)," *University of California Publications in Zoology*, Vol. xxx, pp. 251–392 will be read with interest; and this although the author's numbers are frequently inadequate, and she has not shown whether differences of organs which are not apparently associated with differences of environment may not be correlated with other organs the differences of which are so. Some differences in the skeletal types of different local races of this sparrow she found linked with environment, the greater sternum may be associated with a race which has to take a longer migratory flight, but in others she failed to find any apparent source of differentiation. Because of this failure one has no need to despair of Darwinism!

Breeders of thoroughbreds believe that the increased speed of the modern race-horse is due to careful selection. Only those who have formed pedigrees of thoroughbreds realise how closely inbred they have been, and how the blood of one or two famous horses runs through the whole population. It is not uncommon for a noteworthy stallion to be the sire of 50 to 70 foals in a year, and the amount of inbreeding in thoroughbred horses is amazing. Probably the organs on which speed depends have been pushed nearly to the limit at which they cease to be consistent with the efficiency of other organs. In this respect the continual emphasising of "points" in breeds may lead to their extinction, as it is doing with the bulldog, owing to the high rate of mortality among the littering bitches, which arises from the large heads of the offspring. If we look at wild life there are, I think, various cases where this principle of progressive evolution can be applied with advantage. Its application by no means necessarily involves isolation. We know too little of wild animals to be certain whether they are exogamous or endogamous and to what extent. With man we know that in primitive life exogamy is far from a universal rule and that brother-sister mating has been in the past by no means uncommon. It has been customary in certain tribes for a dead man's wives to fall to his brother, or even to his son. If it has been unlawful at one time and in one place to mate within the clan—except at religious festivals, itself a relic of an early stage of promiscuity within the clan; it has at another time and place been unlawful to mate outside the clan. If such almost instinctive customs and sex-habits hold with man, can we assert that we know enough about animal life to be certain whether exogamy or endogamy is prevalent? What about birds? In 1927 two swallows came and built their nest on a side of my cottage they had not previously frequented. In October they disappeared with their nestlings. In April 1928 five swallows returned and the five birds made straight for the old nest; it seems highly probable that they were the nestlings or the parents and nestlings of the previous year, for the nest must have been familiar to them. Unfortunately they had forgotten that a nest built in the previous year would not support after the winter five fully grown birds and it collapsed under their weight, to their great surprise and alarm. But the incident certainly suggested to me that families of swallows are not broken up in the mass migration flights; and that probably the swallows who built last year round my cottage were all mating either as parents with offspring or brothers with sisters.

Still another case where isolation is unnecessary is that of animals which run in herds. Here the male leads the herd, probably mating with his own offspring until he is supplanted by a stronger and more active male, who will almost certainly be his son. Whether endogamy or exogamy is the customary rule with large carnivores nobody probably knows, but it seems likely that in such cases the members of the litter keep together until they reach the mating stage. Every breeder of mice and dogs knows that such mammals have no repugnancy to endogamy, and that unless the sexes be separated at an early stage, mating may take place within the litter, at what may at first seem a very immature age.

I think therefore that without isolation there may be in nature a fairly wide field for the play of progressive heredity in creating varieties. A good test of this factor might probably be found by investigating whether species which are fairly widely scattered but relatively few in numbers provide many varieties.

Now let us turn to the question of how the principle works in respect of isolation itself. Isolation usually arises in the case of islands, districts separated by broad estuaries, or mountain chains, and in the case of numerous valleys running up into mountains. Even with man the amount of inbreeding in isolated valleys is very little realised. In the Yorkshire Dale from which my father's ancestors came, practically everyone was my great-grandfather's fourth cousin or more nearly related. To-day the great bulk of the inhabitants—although they do not know it—are my sixth cousins. *I* know it from study of the Church registers. Quakerism largely came to an end in the Yorkshire Dale country because of its refusal to sanction cousin marriages. Thus it comes about that in the valleys of Norway you find in one a prevalence of albinism, in another a widespread deformity of the hand, but equally you find valleys in which strength and efficiency are prevalent. It was from such valleys a thousand years ago that the Vikings, men of remarkable vigour, poured down to the devastation or conquest of large parts of Europe.

But to return to lower organic forms, three things in the course of my life have left marked impress on my mind in regard to progressive evolution. In all cases the material was collected by Weldon: (i) *Daphnia* in the ponds round Oxford, each pond appeared to have its own local race, and these races, as his drawings indicated, differ sensibly one from the other; (ii) land snails in the valleys running up to the mountains in Sicily, each valley had a distinct local race of the same species; and (iii) *Draba verna*, here the innumerable sub-races may even be found on opposite sides of the same stone walls.

For Weldon the problem was to find the difference in the chemical constitution of pond-waters, the different rocks, etc. of the valley environments (all carefully photographed), the aspect and stone of the walls which had led to this surprising multiplicity of clearly differentiated local races by, as he thought, *natural selection*. Problems which his short life did not suffice to solve and perhaps could not be solved in a life of any length. As I see the problem now, I think it is not one of the environment and of stringent selection, but of the populating of pond, valley or wall by the offspring of a very limited parentage, which if isolated then inbreed. Powerful as I still hold the influence of natural selection to be, I do not think the differences between local races, or even varieties, are wholly, or in the major part, due to it. Without random mating, but with inbreeding or even some degree of assortative mating only, the various sections of a species will not keep true to type. The basis of differentiation exists in heredity itself, no population is stable, it will break up into castes, unless random mating keeps it to type. Caste-formation in human societies is a natural process of race building, rather than an artificial creation, and on this account may have great evolutionary significance in anthropology.

I think that biologists lay too much stress on the difficulties of procuring isolation. They are rather apt to look upon the world as being in all past time crowded with many individuals of many species. A Surrey common on high ground surrounded by cultivated land is an isolated island for the adders that take refuge there, and probably for many other forms of life as well. The unit of palaeolithic man could hardly have been larger than the family, considering how few are the skeletons we find in the caves he occupied and how few are the caves. It is quite possible that endogamy was then customary. Indeed the fewer the members of a species, the more likely is inbreeding to occur, and the greater will be the profusion of local races.

There is and has been, I hold, a much wider existence of endogamy throughout Nature than some naturalists are inclined to believe; and whether the aversion to exogamy arises from absence of migration in sedentary species, from physiological reasons or from psychical reasons (as in a bitch's distaste for certain dogs), I think there is far more endogamy in wild life than is usually realised.

Granted inbreeding, we should expect as its result that both adverse and favourable characters differing from type would be emphasised. The first thing the inbreeder would note would be heavy mortality from the emphasising of the adverse characters; he would probably overlook the slow if steady development of the favourable characters. This might discourage him from continuing his experiments or drive him to the fatal determination of "introducing new blood." If he persists, notwithstanding the mortality, he would become conscious of the increasing emphasis of certain non-lethal, let us call them for the time being, favourable characters, e.g. increases in certain lines of sizes or weights. Such changes would go on until by the principle of correlation they were checked by corresponding changes in correlated organs, which rendered the organism unsuited to its environment or to its habits. In other words, there would be a mortality rate against extremists, and the survivors would be those of less marked characters.

The above cycle certainly seems to correspond closely to what I have observed generally with my dogs. There was a heavy mortality apparently due, as the breeders would say, to too close inbreeding, but with long continued inbreeding from about five original dogs, three of which were close blood relatives, we have obtained dogs very different from the original types crossed. Further, studying the numerous papers on inbreeding, the above cycle of events seems to be common to all inbreeders' experience, although the explanation usually given is a much more theoretical hypothesis than is needful in the case of pure statistical treatment. Perhaps the most convincing experiments in the sense I have indicated are those on albino rats conducted by Helen Dean King*, who inbred, using brother-sister matings, for 15 generations. The first six generations showed much feebleness which she attributed to malnutrition, but I should attribute to the emphasising of the unfavourable characters. In the seventh generation of inbreeding, she obtained Goliath, the biggest albino rat on record. I exhibit (Plate II) a photograph of him at 18 months. At 15 months he weighed 550 grammes as against stock rats with the average of 320 grammes†. Further, the inbred rats showed greater weight and vigour than the stock rats from the seventh to ninth generation. From the seventh to the fifteenth generation, the inbred male rats were about 18 % heavier than the stock rats. All this is well in accord with what we might expect from the Ancestral Law. I have already indicated to you that, so far as the data for man go, with the ultimate intensity of inbreeding we could not expect a reduction of more than about 30 % in the variability. Dr Helen King says she got the variability reduced by about 40 %. But apart from the fact that Dr King was using a coefficient not proper for measuring a reduction in variation for our present purpose,—she was using the coefficient of variation, while the 30 % reduction is in standard deviation,—we note that in the higher generations at the higher ages she only as a rule measures 1 or 2 male rats. Further, her total number of male rats in each generation is far too small to base any series of numbers upon, e.g.

	Generations								
	7	8	9	10	11	12	13	14	15
Number of rats weighed, 57	3	7	3	5	9	8	6	7	9

---

* *Journal of Experimental Zoology*, Vol. XXVI, 1918: "Studies on Inbreeding, I. The Effects of Inbreeding on the Growth and Variability of Body Weight of the Albino Rat," pp. 1–54. "II. The Effects of Inbreeding on the Fertility and on the Constitutional Vigour of the Albino Rat," pp. 335–378. Vol. XXVII, 1918–19: "III. The Effects of Inbreeding with Selection on the Sex Ratio of the Albino Rat," pp. 1–36. Miss King's papers have been republished as *Studies in Inbreeding* by the Wistar Institute. 1919. 175 pp. Miss King summarises thus her results:

"Comparisons made between inbred and stock animals, reared under similar conditions of environment and of nutrition, show unmistakably that inbred rats are larger, more fertile, and that they attain sexual maturity earlier and possess greater vigor of constitution than do stock animals."

For the applicability of the present theory of progressive evolution, it is, I think, needful to remark that Miss King got rid of the progressively unfavourable deviations by providing that the animals during the inbreeding were of sound stock, i.e. by a "careful selection of the individuals that were used for breeding."

† The standard deviation is about 32 grammes, or the difference is some *seven* times the standard deviation, or say 10 times the probable error.

Fig. (i). Goliath.

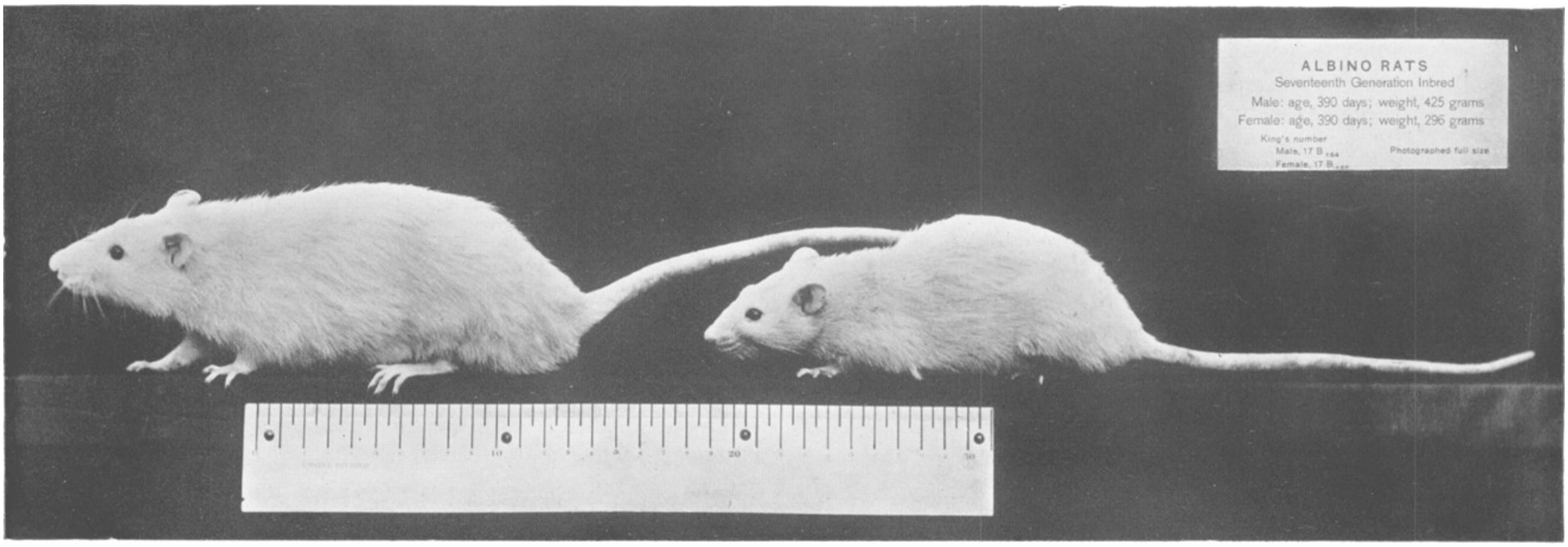


Fig. (ii). 17th Generation of Inbred Rats.

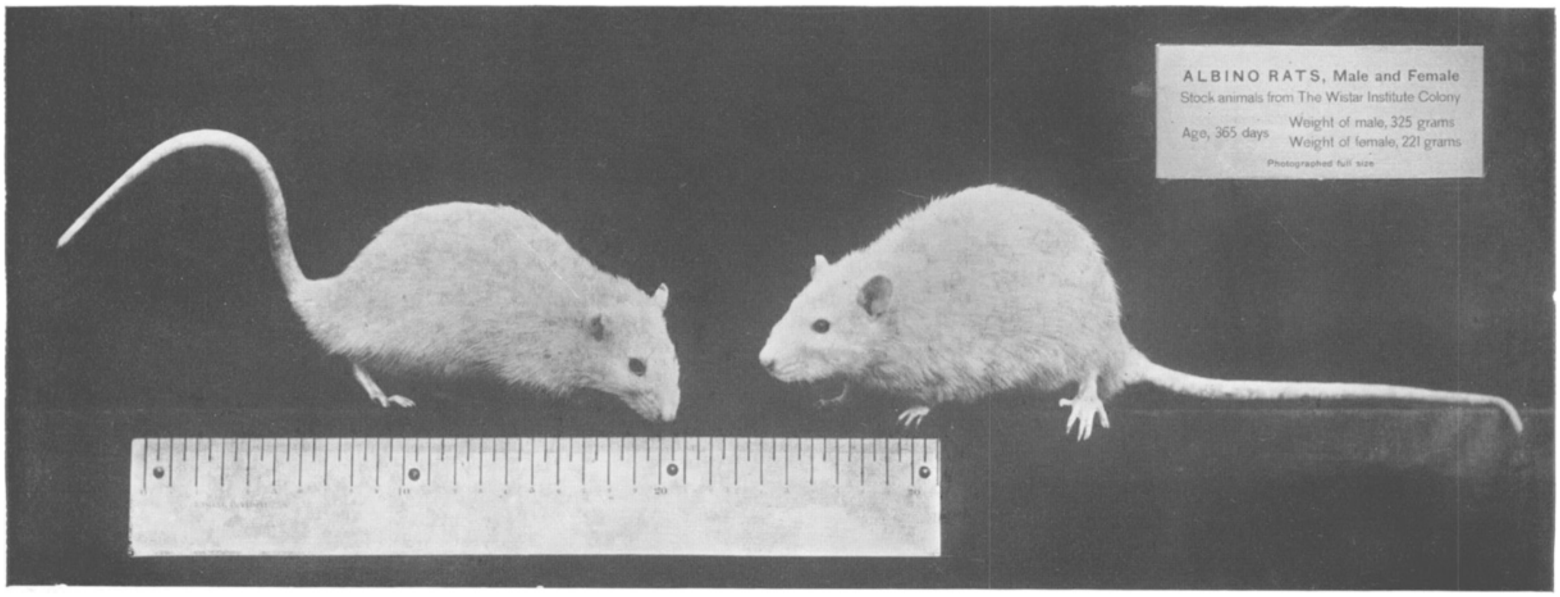


Fig. (iii). Stock Rats for comparison.
Dr Helen Dean King's "Experiments on Inbred Albino Rats."
(*From photographs kindly provided by Dr King.*)

I am quite sure that with such numbers her means and variabilities, still more any percentage decrease in variability, are relatively untrustworthy. However I have done the best I could to obtain results from her figures. I have taken male rats at ages 240 to 245 days. From her Table 17 we find the variabilities as measured by the coefficient of variation are in the fifteenth generation:

Inbred rats, 8·2; Stock rats, 13·9.

From her diagram, Fig. 11, the inbred rats for seventh to fifteenth generation had a mean weight of 315 grammes, her own control a mean weight of 275 grammes and Donaldson's stock rats a mean weight of 268 grammes. We may take the mean weight of a non-inbred rat to be about 270 grammes at 240 days. I have taken 240 days because if we took rats of 15 months Dr King's data for the fifteenth generation would give us one rat to base an average on!

Using the above data, our standard deviations are:

$$\frac{8{\cdot}2 \times 315}{100} = 25{\cdot}83 \text{ for inbred rats,}$$

and

$$\frac{13{\cdot}9 \times 270}{100} = 37{\cdot}53 \text{ for stock rats.}$$

$$\text{Difference} = 11{\cdot}70,$$

and therefore $\frac{11{\cdot}70}{37{\cdot}53} \times 100 = 31\ \%$ reduction by inbreeding.

This is very close to the 30 % about of the human data. I fully admit the precarious nature of the process of reduction. That is not my fault! It arises from the manner in which the biological geneticist appeals to statistics without having any real comprehension of how to handle them. No one without careful examination of Dr King's paper would realise that she is measuring only 57 male rats in all, and that these are spread over 15 generations and over 16 ages, so that her *means* for a given age of a given generation are frequently based on two or even on a single rat record! However, as far as I can see, there is nothing in Dr King's records to make us believe that if you went on inbreeding for still further generations, you could reduce the variability not by 30 % but by 100 %. According to the Mendelians, persistent inbreeding would ultimately produce a number of pure lines, each with zero variability. They account for the facts to which I have been referring, by stating that inbreeding reduces the heterozygosity of the population. If we look at the matter in the simplest case of a two-fold factor, say albinism in man, then we have the homozygous normal men and homozygous albinotic men, and the heterozygous men who carry a factor for albinism. Then if exogamy exists and there be mating at random the population remains statistically stable. Now let inbreeding begin, then the normals produce normals and the albinos albinos; but the heterozygous inbreeding are reduced 50 % at each inbred generation, contributing 25 % to the normals and 25 % to the albinos, who then breed true. Thus the heterozygous rapidly become fewer and fewer and, according to the Mendelians, we should be left in the case of characters depending on multiple factors with a large number of "pure lines." As Castle puts it:

> On a Mendelian theory it would be expected that inbreeding, brother with sister, for a large number of generations would result in a number of homozygous lines, each of which by itself would be entirely devoid of variability, except that due to environmental agencies. (Castle, *Genetics and Eugenics*, 1916.)

Thus the changes observed in inbreeding, as, for example, the increasing weight of rats or the increasing speed of American trotters or of English race-horses, are to be attributed to an elimination of heterozygosity. As there is a gradual increase of the character—in the case of speed of horses possibly for 200 years—the character of speed must depend on a large number of Mendelising units. Thus if we continued inbreeding for a sufficient number of generations, we should ultimately be led to an indefinitely large number of homozygous types; "pure lines," races, even species might be evolved. But from the standpoint of evolution what *does* this mean? If we can obtain from any organism a number of homozygous (true breeding) types by inbreeding, then the alternatives seem to be (i) that the organism was initially heterozygous and local races have arisen by inbreeding, or (ii) that initially every species of organisms consisted of many homozygous lines and that panmixia somehow arose and with it heterozygosity or hybridism.

In the former case there is no explanation, except a direct act of creation for each species, of how any organism originated as a multiple hybrid. Clearly the germ plasm of the organism must then be supposed far more complex initially than it is in cases of much-inbred species.

If, on the other hand, we suppose that initially we had many divergent homozygous races or lines, how did they come into existence? Further, anything we can now obtain by inbreeding must have originally existed as a homozygous race. The high-speed American trotter has existed somewhere before and been swamped by exogamy, until endogamy reproduced him once more!

Thus we seem to face living forms changing in cycles; endogamy alternating with exogamy produces homozygous forms in vast numbers; these a period of exogamy rolls up into a single heterozygous form which will remain stable as long as random mating continues! There is no sign in this alternation of multiple hybridism with innumerable pure lines of any progressive and continuous evolution. If we hold such views, we must be forced to the conclusion, that evolution has had no existence, or if it has had, then it is inexplicable on the Mendelian theory without appeal to mutations. No wonder the prophet of Mendelism declared that he had become "agnostic as to the actual mode and processes of evolution." If the origin of species is to be solved he saw that it did not lie in the Mendelian theory of heredity.

If inbreeding cannot produce anything but homozygous lines, which have somewhere existed before, then there is no progressive evolution. If, however, the pure line be a fiction, and can without hybridisation continue to produce something not itself, then progressive evolution is a possibility whether we are concerned with self-fertilisation or parthenogenesis or merely budding. As you probably all know, the theory of the pure line was started by Johannsen on the basis of the weight of beans (*Phaseolus vulgaris*) bought of a casual seedsman. Anything less characteristic of an individual plant than the weight of one seed produced by that plant it is difficult to imagine! The size of the seed depends on its position in the pod, and the position of the pod on the plant, and it is impossible for the single mother seed to represent the plant or a single daughter seed to do so likewise. But if the pure line theory of Johannsen had been true, then we should expect perfect correlation between mother and daughter plants, and the same between grandparental and daughter plants. By some confusion of mind Johannsen expected them to be and thought he had found these correlations to be *zero*. Actually they were between ·5 and ·6 for the parental and much less for the grandparental relation. How this was to be interpreted Johannsen has never explained.

In 1900 Warren dealt with experiments on the inheritance of characters in *Daphnia* and *Aphis* and found that the grandmaternal correlation was distinctly less than the maternal, a result incompatible with the "pure line" hypothesis. As these results did not suit the views of the supporters of the pure line, they were attributed to change of environment, although Warren had been very careful in his work, using indices as well as absolute measurements. Then came Agar and Hanel, believing they had demonstrated the Johannsen theory by their experiments. Hanel's results were hailed as very definitely confirming Johannsen's hypothesis. When they were examined statistically, however, we were able to point out that they actually gave a lower correlation with grandparent than with parent, and that the parent's correlation with offspring was greater than the parent's fraternity with offspring. The pure linists then cast Hanel on one side, some because they said the number of tentacles in Hydra was not inherited at all, and others with the easy excuse that Hanel, like Warren, had paid no attention to environment!

Next we meet with Jennings, who, experimenting on *Paramecia*, considered he had demonstrated the pure line hypothesis. We have the less need now to criticise his paper, because in more recent work in the case of the rhizopod *Difflugia corona*, he has come to conclusions which are in reality wholly opposed to the pure line theory*. He proceeded to obtain stock from a single progenitor and considered its members with regard to six characters. He says that after many generations of descent from a single progenitor the stock differentiated into many hereditarily diverse stocks. "These diverse stocks differ hereditarily not only with respect to particular single characters, but also with respect to the combinations of characters," so that by selection marked results are obtained. His observations are all the more conclusive because the possible argument as to environment is removed, all the characters dealt with being fixed at fission, and subject to neither growth nor environmental influence. Thus Jennings found selection absolutely effective within the "pure line."

Jennings' paper was published in 1916. In 1919 Hegner obtained similar results for the rhizopod *Arcella dentata*†, producing hereditarily diverse lines from a single progenitor, and claiming that a true case of evolution had been observed in the laboratory. Root‡, Middleton§ and others have produced similar evidence of power to differentiate hereditarily within the "pure line." Perhaps the most ingenious "explanation of this result" is that suggested by Root himself, namely,

> that the pure line concept is correct on the average only; that it is a mathematically correct expression of the mean result of inheritance under natural conditions.

"The pure line hypothesis," we are told, "though true for average results may be thought of as not holding for individual cases." We may be well content with that endeavour to save the hypothesis of Johannsen—the "pure line" is as much a dummy as the electron itself! It does not stand in the way of the non-identity of the various elements of the germ plasm in the same homozygous individual. There is a constant urge to development within the pure line, such as we find in inbred individuals.

In the early days of Mendelism appeals were frequently made to molecular and atomic theories to explain the constancy of the germinal structure. Now-a-days we believe or are supposed to believe that the atom can "waste," or the absolute identity of all atoms of the same species is being replaced by the conception that they can be individual, and the fixity lies in the average or type. If nothing be stable in the physical world except the statistical average, why should we believe that the germinal cells, even in the most homozygous individual, are identical; above all,

---

* "Heredity, Variation and the Results of Selection in the Uniparental Reproduction of *Difflugia corona*," *Genetics*, Vol. I, pp. 407–534, 1916.

† "Heredity, Variation, and the Appearance of Diversities during the Vegetative Reproduction of *Arcella dentata*," *Genetics*, Vol. IV, pp. 95–150, 1919.

‡ "Inheritance in the Asexual Reproduction of *Centropyxis aculata*," *Genetics*, Vol. III, pp. 173–206, 1918.

§ "Heritable Variations and the Results of Selection in the Fision Rate of *Stylonychia pustulata*," *Journal of Experimental Zoology*, Vol. XIX, pp. 450–503, 1915. See also Elmer Roberts in the same *Journal*, "Fluctuations in a Recessive Mendelian Character and Selection," Vol. XXVII, pp. 157–182, 1918–19.

why should we when we know that every organic population, which we can handle, consists of a congeries of differentiated individuals? If the "pure line" hypothesis is to be treated as an average result, why not the Mendelian theory itself? Every known organic population, whatever its process of reproduction, has continuity in its type, diversity in its individuals. Why bind ourselves down to fixity in a population of germ cells? Another strong argument in favour of the non-identical factorial identity of the germ cells at different times may be found in the extraordinary resemblances of the finger-prints of like twins, the greater but lesser likeness of unlike twins and the still less resemblance of ordinary brethren. This is paralleled by the fact that dogs of the same litter from the same bitch and dog have a 30 to 40 % greater resemblance than dogs from different litters of the same parents. These facts are inexplicable on the basis of the absolute identity at all times of the factorial constitution of the germ plasm of the same individual. Every process of inbreeding is a continuous selection of ancestors with divergence from type, and the somatic characters, without being absolutely correlated with the gametic, are sufficiently correlated with the latter to allow of a selection of bodily characters changing the germinal characters. Inbreeding means that we have been selecting the germ plasm for all the characters which differ from type for all the inbred generations. It is a selection of the germ plasm not only of the parents, but, since they in their turn have grandparents of selected germinal constitution, and so on, the selection goes right away back. I submit that the results of inbreeding are not explicable as the disappearance of heterozygosity or as a gradual return to some unspecified homozygous ancestor, who for some occult reason had submitted to hybridisation. They are in themselves evidence of a *continuous* change in type. Individual differences from type will always be emphasised unless the population mates at random. The urge to differentiation, to break into lines, local races, varieties and species, lies in the nature of heredity itself, and is only restrained when there is random mating or panmixia. Isolation and inbreeding lead to differences which ultimately become vital or lethal, and stability is then attained about a new type. While the effects of the urge can certainly be observed in inbreeding mammals, I believe that the clearest demonstrations of progressive evolution will be found in the "pure line" itself, which Johannsen and some of his followers have thought to be unchangeable. But such experiments must be well chosen, above all we must select characters which are certainly inherited and can be demonstrated to be heritable. Otherwise we may be merely asserting the purity of the pure line on the basis of characters which are not inherited or scarcely so at all, as in the case of fertility.

Such is my idea of the progressive urge which heredity itself provides for evolution.

APPENDIX.

The Ancestral Law of Heredity, Collateral Correlations and Prediction Formulae.

It is desirable to remind the reader of the multiple regression formula, and its relation to the ancestral law of heredity.

If there be a system of variates $x_0, x_1 \ldots x_n$, means $\bar{x}_0, \bar{x}_1 \ldots \bar{x}_n$, standard deviations $\sigma_0, \sigma_1 \ldots \sigma_n$, correlation systems given by r_{st} that of x_s and x_t, and D be the determinant

$$\begin{vmatrix} 1, & r_{01}, & r_{02}, & \ldots & r_{0n} \\ r_{10}, & 1, & r_{12}, & \ldots & r_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{n0}, & r_{n1}, & r_{n2}, & \ldots & 1 \end{vmatrix}$$

and D_{st} the minor corresponding to the s, t constituent, then the mean $\tilde{x}_0$ of the array of x_0's for given values of $x_1 \ldots x_n$ is given by

$$\left.\begin{aligned} \tilde{x}_0 - \bar{x}_0 &= \sigma_0 \left\{ - \frac{D_{01}}{D_{00}} \frac{x_1 - \bar{x}_1}{\sigma_1} - \frac{D_{02}}{D_{00}} \frac{x_2 - \bar{x}_2}{\sigma_2} \ldots - \frac{D_{0n}}{D_{00}} \frac{x_n - \bar{x}_n}{\sigma_n} \right\} \\ \text{with a variability} \qquad \Sigma_0 &= \sigma_0 \sqrt{\frac{D}{D_{00}}} = \sigma_0 \sqrt{1 - R^2} \end{aligned}\right\} \quad \ldots\ldots(1),$$

where R is the multiple correlation of x_0 and the best linear function of $x_1, x_2 \ldots x_n$. This result is not dependent upon the normal law of frequency, but to be valuable the regressions among the individual variates must be very approximately linear.

(*a*) *Observational Result. The regression of one relative on a second is very closely linear.* Now if we apply the above regression formula to heredity to predict the *average* character in the offspring, we have an immense number of individual ancestors to deal with, the vast majority of whom can neither have been observed nor their characters recorded. Thus in the pth ascending generation we have 2^p ancestors, and we need some hypotheses to simplify the multiple regression formula. These are as follows:

(*b*) *Degrees of Resemblance the same.* We suppose that on the average the offspring have the same degree of resemblance to each of the ancestors in the same generation*. It accordingly follows that the weight of all the ancestors in the pth generation in the multiple regression formula (1) will be the same, and if we write

$$X_p = \frac{\Sigma_p}{2^p} \left(\frac{{}_p x_1}{{}_p\sigma_1} + \frac{{}_p x_2}{{}_p\sigma_2} + \ldots + \frac{{}_p x_{2^p}}{{}_p\sigma_{2^p}} \right) \quad \ldots\ldots(2),$$

we may rewrite our formula in the form

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(- \frac{\Delta_{01}}{\Delta_{00}} \frac{X_1 - \bar{X}_1}{\Sigma_1} - \frac{\Delta_{02}}{\Delta_{00}} \frac{X_2 - \bar{X}_2}{\Sigma_2} \ldots \right) \quad \ldots\ldots(3),$$

where Δ is the determinant

$$\begin{vmatrix} 1, & \rho_{01}, & \ldots & \rho_{0n} \\ \rho_{10}, & 1, & \ldots & \rho_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{n0}, & \rho_{n1}, & \ldots & 1 \end{vmatrix}$$

and ρ_{st} = correlation of X_s and X_t, the two so-called mid-parents or mid-ancestors in the pth ancestral generation; further,

$$\Sigma_0 = \sigma_0 \sqrt{\frac{\Delta}{\Delta_{00}}} = \sigma_0 \sqrt{1 - R^2},$$

where R is the multiple correlation coefficient. It is then easy to show that

$$\left.\begin{aligned} \Sigma_p &= \frac{1}{2^{\frac{1}{2}p}} \sigma_p \sqrt{1 + \epsilon_p} \\ \rho_{pq} &= \frac{2^{\frac{1}{2}(p-q)} r_{p-q}}{\sqrt{(1 + \epsilon_p)(1 + \epsilon_q)}} \end{aligned}\right\} \quad \ldots\ldots(4),$$

where $p < q$ and ϵ_s is the correlation of husband and wife in the sth generation, i.e. the coefficient of assortative mating. Further, r_{p-q} is the correlation of an individual and any single ancestor in the $(p - q)$th generation above him.

---

* This, except, in very special cases, is of course the assumption made in Mendelian theories, i.e. the sexes are not differentiated.

Now it is clear we could not get further with our prediction formula did we not make some hypothesis as to the change of r_{p-q} with $p - q$. An investigation of a fairly comprehensive Mendelian theory shows that the individual ancestral correlations for somatic characters decrease in a geometrical progression*. This result is in accordance with observed correlations as based on pigmentation records. We therefore make finally the following hypothesis:

(*c*) *The individual ancestral correlations decrease in a geometrical progression.* In other words we assume

$$r_{p-q} = \zeta . \theta^{p-q} \qquad \text{......(5),}$$

and hence, if the coefficient of assortative mating be taken as constant by (4),

$$\rho_{pq} = \frac{\zeta}{1+\epsilon} (\sqrt{2}\theta)^{p-q} = \alpha\beta^{p-q} \qquad \text{......(6),}$$

or the correlations of mid-parents also form a geometrical series.

The next stage in the development of our multiple regression formula is to show that if we write (3) in the form

$$\left. \begin{aligned} \tilde{x}_0 - \bar{x}_0 &= \underset{p=1}{\overset{\infty}{S}} c_p \frac{\sigma_0}{\Sigma_p} (X_p - \bar{X}_p) \\ \Sigma_0 &= \sigma_0 \sqrt{1 - \underset{p=0}{\overset{\infty}{S}} (c_p \rho_{0p})} \end{aligned} \right\} \qquad \text{......(7),}$$

then

$$c_p = \gamma\eta^p,$$

or the coefficients of the multiple regression formula themselves form a geometrical series.

The determination of γ and η follows from the equations

$$\eta^2 - \eta \frac{1 - 2\alpha\beta^2 + \beta^2}{\beta(1-\alpha)} + 1 = 0,$$

where we must take the root of η less than unity, and

$$\gamma = (\beta - \eta)/\eta \qquad \text{......(8).}$$

Equation (5) shows us that if the correlation of parent and offspring, and grandparent and offspring, be known, then ζ and θ are determined. Hence by (6)

$$\alpha = \zeta/(1+\epsilon), \quad \beta = \sqrt{2}\theta \qquad \text{......(9)}$$

are known, or the mid-parental correlations are ascertained. We thus have from (7)

$$\left. \begin{aligned} \tilde{x}_0 - \bar{x}_0 &= \underset{p=1}{\overset{\infty}{S}} \gamma\eta^p \frac{\sigma_0}{\Sigma_p} (X_p - \bar{X}_p) \\ \Sigma_0 &= \sigma_0 \sqrt{1 - \frac{\gamma\eta\alpha\beta}{1-\eta\beta}} \\ \text{and} \quad R^2 &= \frac{\gamma\eta\alpha\beta}{1-\eta\beta} \end{aligned} \right\} \qquad \text{......(10).}$$

These expressions form what I term the Ancestral Law of Heredity, it is a prediction formula for offspring given the complete ancestry. It depends solely on the hypotheses and observational results (*a*), (*b*) and (*c*), and it suffices to find *two* individual ancestral correlations to determine the whole chain of resemblances to direct and collateral ancestry of the offspring. I term the expression

$$g = \underset{p=1}{\overset{\infty}{S'}} \gamma\eta^p \frac{\sigma_0}{\Sigma_p} (X_p - \bar{X}_p),$$

---

* *Phil. Trans.* Vol. 203 A, pp. 53–86, 1904.

the *generant* for a given character of a given individual. The generant of an individual is therefore that linear function of the same characters in his entire ancestry, which has the maximum correlation (R) with the character in the individual. Two brothers have the same generant, but two cousins have only in part the same generant. I term such common part generants *partial generants.* The correlation of collaterals depends on separating out their partial generants. If we take ${}_1x_0$, ${}_2x_0$ as the character values in two brothers, then

$$ {}_1x_0 = \tilde{x}_0 + {}_1x_0' = \bar{x}_0 + {}_1x_0' + g, $$

$$ {}_2x_0 = \tilde{x}_0 + {}_2x_0' = \bar{x}_0 + {}_2x_0' + g, $$

and

$$ ({}_1x_0 - \bar{x}_0)\,({}_2x_0 - \bar{x}_0) = {}_1x_0' \times {}_2x_0' + g\,({}_1x_0' + {}_2x_0') + g^2. $$

If curled brackets denote mean values when we sum for two brothers out of every array, and r be the correlation of brothers in the ordinary sense,

$$ r = \frac{\{({}_1x_0 - \bar{x}_0)\,({}_2x_0 - \bar{x}_0)\}}{\sigma_0^2} = \frac{\{g^2\}}{\sigma_0^2}, $$

for ${}_1x_0'$ and ${}_2x_0'$ are uncorrelated and g is not correlated with ${}_1x_0'$ or ${}_2x_0'$, these being the variations within the array of two brothers selected at random.

Accordingly we have to find $\{g^2\}/\sigma_0^2$. Returning to the notation of (7) this is easily seen to be equal to

$$ \mathop{S}\limits_{p=1}^{\infty} (c_p^2 + 2c_p c_{p'} \rho_{p-p'}) $$

or

$$ \frac{\{g^2\}}{\sigma_0^2} = c_1\rho_{01} + c_2\rho_{02} + \dots $$

$$ = \frac{\gamma\eta\alpha\beta}{1 - \eta\beta} = R^2 \qquad \text{......(11).} $$

Thus we have the correlation of brothers equal to the squared multiple correlation of a man with all his ancestry. This result, obtained more directly, is well known*. Since $r = R^2$ it is clear that all the constants of the ancestral law can be found from the parental and fraternal individual correlations as well as from the parental and grandparental individual correlations.

The value of β is

$$ \left.\begin{aligned} \beta &= \frac{r_1\sqrt{2}}{1+\epsilon} + \sqrt{(1 - r)\left(1 - \frac{2r_1^2}{(1+\epsilon)^2\,r}\right)} \\ \text{and}\qquad \alpha\beta &= \frac{r_1\sqrt{2}}{1+\epsilon} \end{aligned}\right\} \qquad \text{......(12).} $$

Thus α and β are known, when we have: (a) the parental correlation r_1, (b) the fraternal correlation r, and (c) the coefficient of assortative mating ϵ.

Since for all values at present worked out α is = or < 1, and β must be positive and = or < 1, it follows that the plus sign must be given to the radical. α and β being known, (9) gives ζ and θ, and thus all the individual ancestral correlations. Lastly:

$$ \eta = \frac{r - \alpha\beta^2}{\beta\,(r - \alpha)} \quad \text{and} \quad \gamma = \frac{\beta - \eta}{\eta} \qquad \text{......(13)} $$

---

* *Biometrika*, Vol. XVII, p. 131.

complete the solution as determined from the parental and fraternal correlations. The equation to determine β given r_1, r and ϵ is really a biquadratic, but the factor $\beta^2 - 1$ divides out. Thus a possible root of it is $\beta = 1$. In this case the process by which we have solved our infinite equations is no longer valid, but $\beta = 1$ has a very valid genetic significance. It denotes that all the ρ's are equal, or all the mid-ancestors are equally correlated with the offspring and with each other. Each generation of ancestors contributes equally to the prediction and all generations are equivalent as a means of testing the probable offspring. A knowledge of the somatic characters of eight great-grandparents is as serviceable as that of four grandparents or of two parents. It follows at once that in Equation (7) the regression constants c_p must all be equal.

To predict therefore from n ancestral generations is in this special case equivalent to predicting the character of one from the remaining n out of $n + 1$ equally correlated individuals. That is, we return to 1 and evaluate the determinant of $n + 1$ columns and rows

$$D = \begin{vmatrix} 1, & \alpha, & \alpha, & \ldots & \alpha \\ \alpha, & 1, & \alpha, & \ldots & \alpha \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \alpha, & \alpha, & \alpha, & \ldots & 1 \end{vmatrix}.$$

and its minors.

Here, by (9), $\theta = \frac{1}{\sqrt{2}}$ and $\alpha = r_1 \sqrt{2}/(1 + \epsilon)$.

We obtain the following formulae:

$$\left.\begin{aligned} \tilde{x}_0 - \bar{x}_0 &= \overset{n}{\underset{p=1}{S}} \frac{\sigma_0 \alpha}{1 + (n-1)\alpha} \frac{X_p - \bar{X}_p}{\Sigma_p} \\ \Sigma_0 &= \sigma_0 \sqrt{1 - R^2} = \sigma_0 \sqrt{\frac{(1-\alpha)(1+n\alpha)}{1+(n-1)\alpha}} \\ \text{and} \qquad R^2 &= \frac{n\alpha^2}{1 + (n-1)\alpha} \end{aligned}\right\} \qquad \ldots\ldots(14).$$

Now when n is indefinitely great $R^2 = \alpha$, hence α is the correlation of brothers $= r$. It is a condition therefore of $\beta = 1$, that

$$r = r_1 \sqrt{2}/(1 + \epsilon) \qquad \ldots\ldots(14^{\text{bis}}).$$

Actually for stature, span and cubit in man the mean value of the fraternal correlation in the nine tables is ·519, and the mean value of $r_1 \sqrt{2}/(1 + \epsilon)$ in the corresponding twelve tables is ·529; thus we may reasonably anticipate that β is not far from unity for man.

We will examine these results in their bearing on a single character, stature, in man. Here the mean of the four parental tables is a correlation of ·50625 and of the three fraternal tables of ·53367, while the coefficient of assortative mating in stature is ·2804. Accordingly our constants are

$$r_1 = \cdot 50625, \quad r = \cdot 53367, \quad \epsilon = \cdot 2804.$$

Hence we find $\qquad r_1 \sqrt{2}/(1 + \epsilon) = \cdot 55916,$

and accordingly by (12) $\qquad \beta = 1 \cdot 0014.$

Thus β is essentially unity and thus $\qquad \alpha = \cdot 55916,$

which is the mutual correlation of all mid-ancestors with one another and the offspring.

In a previous investigation I separated the stature for sons and daughters and found the β for the former ·99976 and for the latter ·99902; the respective α's were ·5567 and ·5616, giving a mean value ·55915 in excellent agreement with the above.

The correlation of the pth generation individual ancestor with the offspring is

$$\begin{aligned} r_p &= \zeta\theta^p = \alpha(1+\epsilon)(\beta/\sqrt{2})^p = r_1\sqrt{2}\left(\frac{1}{\sqrt{2}}\right)^p \\ &= \cdot 50625\,(\cdot 707{,}107)^{p-1}. \end{aligned}$$

Our series will then be as follows*:

Parent and Offspring	·5063
Grandparent and Offspring	·3580
Great Grandparent and Offspring	·2531
Great Great Grandparent and Offspring	·1790
Great Great Great Grandparent and Offspring ...	·1266
Great Great Great Great Grandparent and Offspring	·0895

Under the circumstances of $\beta = 1$, the law of ancestral heredity now takes the form for prediction from n generations,

$$\left.\begin{aligned} \tilde{x}_0 - \bar{x}_0 &= \frac{\cdot 55916\sigma_0}{1+(n-1)\cdot 55916}\overset{n}{\underset{p=1}{S}}\frac{X_p - \bar{X}_p}{\Sigma_p} \\ \Sigma_0 &= \sigma_0\sqrt{\frac{\cdot 44084\,(1+n\cdot 55916)}{1+(n-1)\cdot 55916}} \end{aligned}\right\} \quad \ldots\ldots(15).$$

Or, we see that while continual selection will raise the type indefinitely, yet with the amount of assortative mating we find in man, we cannot hope even by the longest continued selection to reduce the variability by more than about 34 %.

Returning to Equation (4) we can substitute for Σ_p and read the first equation of (15) as

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = \frac{\cdot 494{,}155}{1+\cdot 55916\,(n-1)}\overset{n}{\underset{p=1}{S}}\left((\sqrt{2})^p\,\frac{X_p - \bar{X}_p}{\sigma_p}\right) \quad \ldots\ldots(16).$$

Now let us suppose that the population has been mating at random and that we then select two parents, whose deviations from the population means are h times their standard deviations†. Then (i)

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = \frac{\cdot 494{,}155\times\sqrt{2}}{1+\cdot 55916\times 0}h = \cdot 698{,}841h \;(F_1 \text{ generation}).$$

This is the real regression of offspring on parents, first noted by Francis Galton.

(ii) We now inbreed so that the deviation of the parents is ·698,841h × their standard deviation, but that of the grandparents will be h × their standard deviation. Here

$$\begin{aligned} \frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} &= \frac{\cdot 494{,}155}{1\cdot 55916}(\sqrt{2}\times\cdot 698{,}841 + 2)\,h \\ &= \cdot 947{,}105h \;(F_2 \text{ generation}). \end{aligned}$$

Still some regression on grandparents, but a progression on the parents.

---

* Cf. *Biometrika*, Vol. XVII, p. 139.

† We are supposing female deviations reduced to male equivalents by using the ratio of male to female standard deviations.

(iii) We now take three generations of inbreeding, corresponding to parents ·947,105*h*, grandparents ·698,841*h* and great grandparents *h*. We have now

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = \frac{\cdot 494{,}155}{2\cdot 11832} (\sqrt{2} \times \cdot 947{,}105 + 2 \times \cdot 698{,}841 + 2\sqrt{2})\, h$$

$$= 1\cdot 298{,}306 h \ (F_3 \text{ generation}),$$

that is to say with three generations of inbreeding the regression has become a progression and the offspring have a 30 % greater deviation from the racial mean.

(iv) For the F_4 inbred generation we have in the same manner

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = 1\cdot 791{,}505 h.$$

(v) For the F_5 inbred generation we have

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = 2\cdot 482{,}687 h.$$

(vi) For the F_6 inbred generation there results

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = 3\cdot 450{,}921 h.$$

(vii) For the F_7 inbred generation there results

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = 4\cdot 807{,}492 h.$$

(viii) And lastly, for the F_8 or eighth inbred generation,

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = 6\cdot 708{,}885 h.$$

Thus we see with each successive inbred generation from the third the type of the offspring $\tilde{x}_0$ continues progressively to diverge from the racial type $\bar{x}_0$, and this will be true whether *h* be positive or negative, and whether the deviation be above or below the mean; and however small the excess or defect may be it will be ultimately magnified into a sensible magnitude if the inbreeding be continued. Ultimately the character under discussion, or one or another correlated character, may have evolutionary importance whether for its survival or lethal value. The Law of Ancestral Heredity does not of course explain the mechanism of the process by which inbreeding will lead to a continuous or progressive deviation in the resulting offspring from racial type. But the well-substantiated values for parental and fraternal correlation for stature in man form a strong argument that there are characters such that inbreeding will lead to a progressive change in the absence of natural selection.

It will be seen at once that the view here expressed is absolutely opposed to that of Johannsen. If by "pure line" we signify one in which inbreeding only occurs, then according to the above view the individual pure line will continue to progress in its own particular direction and not remain stable as Johannsen asserted. Stability depends upon random mating, and assortative mating with its limit pure inbreeding connotes continuous change.

Returning to Equation (10), we may write it in the form

$$\frac{\tilde{x}_0 - \bar{x}_0}{\sigma_0} = \overset{\infty}{\underset{p=1}{S}} \frac{1}{\sqrt{1+\epsilon}} \gamma \, (\eta \sqrt{2})^p \frac{X_p - \bar{X}_p}{\sigma_p},$$

and accordingly, if we were to suppose that no amount of selection of ancestors of deviation $h\sigma_p$ could produce more than a deviation $h\sigma_0$ in the ancestry, it would follow that we must have

$$\frac{1}{\sqrt{1+\epsilon}}\,\frac{\gamma\eta\sqrt{2}}{1-\eta\sqrt{2}} = 1,$$

or, by the second equation of (8),

$$\beta\sqrt{2} - \eta\sqrt{2} = \sqrt{1+\epsilon}\,(1 - \eta\sqrt{2}) \qquad \text{......(17).}$$

This is the relation under the circumstances of β to η. If there be random mating, $\epsilon = 0$, and $\beta = 1/\sqrt{2}$. This leads by (9) to $\theta = \frac{1}{2}$ and $\zeta = \alpha$. Hence since $\zeta\theta = r_1$, the first individual parental correlation, we have a series of individual ancestral correlations decreasing by the radix $\frac{1}{2}$.

Simple illustrations of such stable populations, where mating at random is the rule and where definite selection of the same extent of deviation for however many generations will not lead the offspring type beyond the ancestral character, are:

(i) Galton's Law of Ancestral Contributions.

Here $\alpha = \cdot 6$, $\beta = 1/\sqrt{2}$, $r_1 = \cdot 3$ and the pth individual ancestral correlation $r_p = \cdot 3 \times (\frac{1}{2})^{p-1}$.

(ii) Simple Mendelian allelomorphs. Somatic characters.

Here $\alpha = \frac{2}{3}$, $\beta = 1/\sqrt{2}$, $r_1 = \frac{1}{3}$ and $r_p = \frac{1}{3}\,(\frac{1}{2})^{p-1}$.

(iii) Simple Mendelian allelomorphs. Gametic characters.

Here $\alpha = 1$, $\beta = 1/\sqrt{2}$, $r_1 = \frac{1}{2}$ and $r_p = (\frac{1}{2}) \times (\frac{1}{2})^{p-1}$ and we may note $\eta = 0$, while $\gamma\eta = \beta = \frac{1}{2}$.

All such populations exclude progression under random mating conditions, and with inbreeding do not lead to a type of offspring exceeding in character that of the selected ancestry.

Hitherto we have considered what happens when we inbreed, or take the offspring of selected ancestors, but if we form prediction formulae for a sufficient number of relations other than ancestors, we soon find the type of the offspring group* or sibship exceeding in the character under consideration that of the selected ancestry, if the parental and fraternal correlations have values similar to those determined in the case of man. I do not think I have seen regression formulae worked out for various groups of selected relatives and I propose to give a few here, using Equation (1) to determine the fitting formula. As I have indicated, all the correlations for individual pairs of relatives can be found from a knowledge of r_1 and r the parental and fraternal correlations, and I shall proceed to determine some of these before discussing the multiple regression formulae for the types in sibships related to specially selected sibs.

But before we do this the reader may inquire whether there exists no general relation between r_1 and r, the parental and fraternal correlations. I do not think this is so, unless we make some additional hypothesis. Clearly to find θ, the rate at which individual ancestral correlations decrease, we must know β, for by (6) $\theta = \beta/\sqrt{2}$. Now β is given by (12), which involves both r and r_1, as well as ϵ. If we suppose as our hypothesis $\beta = 1$, which is very nearly true for measurable characters in man, then (15) holds and r is determined from r_1 and ϵ, and this reduces to $r = r_1\sqrt{2}$, if there be no assortative mating. Such a result is very much out in the case of man, where ϵ is certainly not zero, varying from $\cdot 17$ to $\cdot 28$. A second hypothesis to link r with r_1 lies in supposing the population stable. In this case (17) holds; and if we superimpose the condition of random mating $\beta = 1/\sqrt{2}$ by (12),

$$\frac{1}{\sqrt{2}} - r_1\sqrt{2} = \sqrt{(1-r)\left(1 - \frac{2r_1^2}{r}\right)},$$

or

$$2r^2 - r\,(1 + 4r_1) + 4r_1^2 = 0.$$

---

* This term is incorrect for the group, the type of which we are studying, is not necessarily the progeny of all or any of the selected relations. It is better to term this array the *sibship* from the given relatives or selected *sibs*.

This may be put into the form $r(r - {\cdot}5) = 2r_1(r - r_1)$.

Thus since r and r_1 will always be positive we see that for $r < {\cdot}5$, r_1 will be $> r$ or parental correlation greater than fraternal. When

$$\begin{aligned} r_1 &= {\cdot}55 \text{ then } r = {\cdot}6129, \\ r_1 &= {\cdot}50 \quad ,, \quad r = {\cdot}5000, \\ r_1 &= {\cdot}45 \quad ,, \quad r = {\cdot}4085, \\ r_1 &= {\cdot}40 \quad ,, \quad r = {\cdot}3298, \\ r_1 &= {\cdot}35 \quad ,, \quad r = {\cdot}2397, \\ r_1 &= {\cdot}30 \quad ,, \quad r = {\cdot}2000. \end{aligned}$$

Now parental correlations are very rarely above ·50, and fraternal correlations are most often somewhat in excess of parental, hence very little can be said for hypotheses like the above which suppose random mating and a stable population. I doubt very much whether any general relation between r and r_1 is discoverable, which would be of practical service, at any rate in the case of measurable characters in man.

I now propose to determine, from a knowledge of r and r_1, the correlations between first cousins, $r_{c_1 c_1}$, between second cousins, $r_{c_2 c_2}$, between uncle and nephew, $r_{u_1 n}$, and between great uncle and nephew, $r_{u_2 n}$. These correlations will be discoverable by considering the partial generants which these relatives have in common. A word of warning is here desirable; there is no doubt of the existence of assortative mating in man, there is a correlation between husband and wife, significant if not of a high order. It is conceivable, accordingly, that a correlation may exist between a man and members of his wife's stirp. Since the correlation of assortative mating, although significant, is small, this correlation, even in the case of a man and his wife's sister, father or mother, is likely to be smaller still, a second order correlation. As far as I am aware such correlations have not been investigated, and while considering the correlation of husband and wife, we shall suppose the relation between the generants of a man and of his wife to be negligible*.

Taking first cousins, since we suppose all individual correlations independent of sex, we may take the sons of two brothers to represent our cousins. These two brothers will have the same generant g_1, but their two wives will have generants g_2 and g_3 which have no correlation one with the other or with g_1. We have accordingly to break up the total generant of a man into the generants partial to his father and mother.

We will first express the character x_0 of an individual in terms of his first order partial generants g_1 and g_2, i.e. the generants of his father F_1 and his mother M_1; we suppose him to have a deviation x_0' from the mean value $\tilde{x}_0$ of his fraternal array. Then it is easy to show that

$$x_0 - \bar{x}_0 = x_0' + g = x_0' + \frac{\gamma\eta}{\sqrt{2(1+\epsilon)}}\left(\frac{\sigma_0}{\sigma_1}(F_1 - \bar{F}) + \frac{\sigma_0}{\sigma_1'}(M_1 - \overline{M})\right) + \frac{\eta}{\sqrt{2}}(g_1 + g_2) \quad \ldots(18).$$

Now we know that $\{g^2\}/\sigma_0^2 = R^2$, by (11), and we shall require $\{(x_0 - \bar{x}_0)\,g\}/\sigma_0^2$. The first thought that might occur would be that we have here the correlation of a man with his generant, but this is not so. The correlation of a man and his generant is

$$R = \{(x_0 - \bar{x}_0)\,g\}/(\sigma_0\sigma_g),$$

and we have $\qquad \sigma_g^2 = \{g^2\} = R^2\sigma_0^2;$

thus it follows that $\qquad \{(x_0 - \bar{x}_0)\,g\}/\sigma_0^2 = R^2$, and not R $\qquad$ ......(19).

---

* We assume of course that the man and his wife have no consanguinity.

We are now in a position to determine the correlation of cousins. Their generant formulae are

$$ {}_{c}x_0 - \bar{x}_0 = {}_{c}x_0' + \frac{\gamma\eta}{\sqrt{2\,(1+\epsilon)}}\left(\frac{\sigma_0}{\sigma_1}(F_1 - \bar{F}) + \frac{\sigma_0}{\sigma_1'}(M_1 - \bar{M})\right) + \frac{\eta}{\sqrt{2}}(g_1 + g_2), $$

$$ {}_{c'}x_0 - \bar{x}_0 = {}_{c'}x_0' + \frac{\gamma\eta}{\sqrt{2\,(1+\epsilon)}}\left(\frac{\sigma_0}{\sigma_1}(F_1' - \bar{F}) + \frac{\sigma_0}{\sigma_1'}(M_1' - \bar{M})\right) + \frac{\eta}{\sqrt{2}}(g_1 + g_2'). $$

To obtain the correlation of cousins we must multiply these two expressions together and take their mean value for the whole population. Now ${}_{c}x_0'$ and ${}_{c'}x_0'$ are random deviations within the fraternal arrays corresponding to the two cousins and have no correlations with each other, with the individual fathers and mothers or with the partial generants. F_1 and F_1' are brothers, but neither has any relation to their brother's wife's generants. Accordingly,

$$ \{({}_{c}x_0 - \bar{x}_0)\,({}_{c'}x_0 - \bar{x}_0)\} = r_{cc'}\sigma_0^2 = \frac{\gamma^2\eta^2}{2\,(1+\epsilon)}\sigma_0^2 r + \frac{\eta^2}{2}\{g_1^2\} + \frac{\gamma\eta^2\sigma_0^2}{2\sqrt{1+\epsilon}}\left\{\frac{g_1}{\sigma_0}\frac{(F_1 - \bar{F})}{\sigma_1}\right\} + \frac{\gamma\eta^2\sigma_0^2}{2\sqrt{1+\epsilon}}\left\{\frac{g_1}{\sigma_0}\frac{F_1' - \bar{F}}{\sigma_1}\right\}, $$

and hence, by (11) and (19), we have

$$ r_{cc'} = \frac{1}{2}\frac{\eta^2}{1+\epsilon}(\gamma^2 r + 2\gamma\sqrt{1+\epsilon}\,R^2 + (1+\epsilon)\,R^2), $$

or, since $R^2 = r$,

$$ r_{cc'} = \tfrac{1}{2} r\eta^2\left(1 + \frac{\gamma}{\sqrt{1+\epsilon}}\right)^2 \qquad \text{......(20).} $$

This is the correlation of cousins, and it is expressible in terms of the fraternal and parental correlations and the coefficient of assortative mating by the equations we have already provided. To determine the avuncular correlation all we need is to multiply (18) by $(F_1' - \bar{F})/\sigma_1$, where F_1' is the brother of F_1 and uncle of x_0, and take mean values. We find

$$ \left\{(x_0 - \bar{x}_0)\frac{(F_1' - \bar{F})}{\sigma_1}\right\} = r_{nu}\sigma_0 = \frac{\gamma\eta}{\sqrt{2\,(1+\epsilon)}}\sigma_0 r + \frac{\eta\sigma_1}{\sqrt{2}}\left\{\frac{F_1' - \bar{F}}{\sigma_1}\frac{g_1}{\sigma_1}\right\} $$

$$ = \frac{\gamma\eta\sigma_0}{\sqrt{2\,(1+\epsilon)}}r + \frac{\eta}{\sqrt{2}}\sigma_1 R^2 \text{ by (19),} $$

or

$$ r_{nu} = \frac{\eta r}{\sqrt{2}}\left(\frac{\sigma_1}{\sigma_0} + \frac{\gamma}{\sqrt{1+\epsilon}}\right) \qquad \text{......(21).} $$

If we may treat the variability as constant for two successive generations, $\sigma_1/\sigma_0 = 1$, and

$$ r_{nu} = r\frac{\eta}{\sqrt{2}}\left(1 + \frac{\gamma}{\sqrt{1+\epsilon}}\right) \qquad \text{......(22).} $$

I owe to Miss Elizabeth Worthington the following further expressions for collateral correlations obtained by the use of partial generants of higher order:

Correlation of second cousins c_2 and c_2',

$$ r_{c_2c_2'} = r\frac{\eta^4}{4}\left(1 + \frac{\gamma}{\sqrt{1+\epsilon}}\right)^4 \qquad \text{......(23).} $$

Correlation of nephew n and great uncle u_2,

$$ r_{nu_2} = r\frac{\eta^2}{2}\left(1 + \frac{\gamma}{\sqrt{1+\epsilon}}\right)^2 \qquad \text{......(24).} $$

Thus the correlation of nephew and great uncle is the same as that of first cousins.

It is clear that the expression

$$\lambda = \frac{\eta}{\sqrt{2}}\left(1 + \frac{\gamma}{\sqrt{1+\epsilon}}\right) \qquad \text{......(25)}$$

plays an important part in the heredity factor between collaterals. It may be termed the "collateral colligator" or simply the "colligator."

In reckoning nearness of relationship in the matter of collaterals, we must count up the number of generations until the two collaterals have ancestors who were brothers, and the sum of these generations for the two collaterals gives the power to which the colligator must be raised before it is multiplied into the fraternal correlation r. Thus if collateral a has p generations to ascend before he reaches the individual A who was the brother of B from whom collateral b is descended by q generations, the correlation of a and b is

$$r_{ab} = r_{AB}\,\lambda^{p+q} = r\lambda^{p+q}.$$

Thus we have

Correlation of Brothers*	$p = q = 0,\ r_{ab} = r$
Correlation of Nephew and Uncle	$p = 1,\ q = 0,\ r_{ab} = r\lambda$
Correlation of First Cousins	$p = 1,\ q = 1,\ r_{ab} = r\lambda^2$
Correlation of Nephew and Great Uncle	$p = 2,\ q = 0,\ r_{ab} = r\lambda^2$
Correlation of Second Cousins	$p = 2,\ q = 2,\ r_{ab} = r\lambda^4$
Correlation of a Man and his First Cousin once removed	$p = 1,\ q = 2,\ r_{ab} = r\lambda^3$

and so on.

Francis Galton argued that a nephew (n) was the son of a father's brother (u)† and made the correlation $r_{nu} = r \times r_1$, where r_1 is the parental correlation. He thus equated the colligator λ to the parental correlation r_1. I do not think this was legitimate.

From Equations (8) and (10) we know that

$$\gamma\eta = \beta - \eta \quad\text{and}\quad r = \gamma\eta\alpha\beta/(1 - \eta\beta),$$

---

* *Half-Brothers* $\frac{1}{2}B$, $\frac{1}{2}B'$. It may not be without interest to record the correlation of half-brothers. If ${}_1x_0$, ${}_2x_0$ be the characters in two half-brothers, we have

$${}_1x_0 - \bar{x}_0 = {}_1x_0' + \frac{\gamma\eta}{\sqrt{2(1+\epsilon)}}\left(\frac{\sigma_0}{\sigma_1}(F_1 - \bar{F}) + \frac{\sigma_0}{\sigma_1'}(M_1 - \bar{M})\right) + \frac{\eta}{\sqrt{2}}(g_1 + g_2),$$

$${}_2x_0 - \bar{x}_0 = {}_2x_0' + \frac{\gamma\eta}{\sqrt{2(1+\epsilon)}}\left(\frac{\sigma_0}{\sigma_1}(F_1 - \bar{F}) + \frac{\sigma_0}{\sigma_1'}(M_2 - \bar{M})\right) + \frac{\eta}{\sqrt{2}}(g_1 + g_2').$$

The father F_1 has a correlation ϵ with both mothers but is assumed to have none with g_2 and g_2'. M_1 is independent of g_1 and g_2' and M_2 of g_1 and g_2. Multiply the two expressions together and take the mean for the whole population, and we have

$$\{({}_1x_0 - \bar{x}_0)({}_2x_0 - \bar{x}_0)\} = r_{\frac{1}{2}b\,\frac{1}{2}b'}\,\sigma_0^2 = \sigma_0^2\left(\frac{\gamma^2\eta^2}{2(1+\epsilon)}(1 + 2\epsilon) + \frac{\gamma\eta^2}{1+\epsilon}\left\{\frac{g_1}{\sigma_0}\,\frac{F_1 - \bar{F}}{\sigma_1}\right\} + \tfrac{1}{2}\eta^2\frac{\{g_1^2\}}{\sigma_0^2}\right).$$

Hence by (11) and (19),

$$r_{\frac{1}{2}b\,\frac{1}{2}b'} = \frac{\gamma^2\eta^2(1 - r + 2\epsilon)}{2(1+\epsilon)} + r\frac{\eta^2}{2}\left(1 + \frac{\gamma}{\sqrt{1+\epsilon}}\right)^2 = r_{cc'} + \frac{\gamma^2\eta^2(1 - r + 2\epsilon)}{2(1+\epsilon)} \qquad \text{......(25) bis,}$$

where $r_{cc'}$ is the correlation of first cousins. The correlation of half-brothers is therefore slightly greater (in man) than the correlation of first cousins. Substituting the values for stature we find, numerically,

$$r_{\frac{1}{2}b\,\frac{1}{2}b'} = \cdot 2474 + \cdot 0045 = \cdot 2519.$$

We may note that half-brothers have very nearly half the correlation ($\frac{1}{2}\cdot 5110 = \cdot 2555$) of brothers.

† See *Natural Inheritance*, p. 133, and my *Life of Francis Galton*, Vol. III, p. 24.

and these allow us to find η and $\gamma\eta$ in terms of α, β and r. We have

$$\eta = (r - \alpha\beta^2)/(\beta r - \alpha\beta), \quad \gamma\eta = r(1 - \beta^2)/(\beta r - \alpha\beta),$$

but by (12), $\alpha\beta = r_1\sqrt{2}/(1 + \epsilon)$, and accordingly

$$\lambda = \frac{1}{\sqrt{2}}\left(\beta\frac{r_1\sqrt{2}}{1+\epsilon} - r - \frac{r(1-\beta^2)}{\sqrt{1+\epsilon}}\right)\frac{1}{\dfrac{r_1\sqrt{2}}{1+\epsilon} - \beta r} \qquad \text{......(26).}$$

Clearly when, as in man, β approaches unity, λ approaches $1/\sqrt{2}$ or ·7071. The colligator is thus considerably greater than the parental correlation, which is rarely in excess of ·5000.

In any other case it seems needful to find β from (12) and insert its value in (26).

I take the following example for stature, confining myself to sons,

$$\epsilon = \cdot 2804, \quad r_1 = \cdot 5040^*, \quad r = \cdot 5110.$$

Whence we deduce from (12) $\beta = \cdot 995{,}372$†,

and from (26) $\lambda = \frac{1}{\sqrt{2}} \times \cdot 983{,}9502 = \cdot 695{,}758$.

Thus β is nearly unity and λ nearly $1/\sqrt{2}$. Accordingly we have

Correlation of Uncle and Nephew	·3555 $[r_\lambda]$
Correlation of First Cousins	·2474 $[r_{\lambda^2}]$
Correlation of Great Uncle and Nephew	
Correlation of Man and First Cousin once removed	·1721 $[r_{\lambda^3}]$
Correlation of Second Cousins	·1197 $[r_{\lambda^4}]$

Thus a knowledge of λ and r provides us at once with all degrees of collateral relationship. It is clear that these degrees of blood relationship are not without interest in the case of men dying intestate without leaving widow or offspring. Thus a nephew has more claim, as nearer of blood than a cousin; a first cousin once removed than a second cousin, and a great nephew and a first cousin equal claims. These grades of proximity may have been generally recognised, but probably hitherto no numerical values have been attached to them.

A very simple notation for the correlation of collaterals can now be introduced; for the nth degree of affinity the correlation is

$$r_{\lambda^n} = r\lambda^n,$$

and we shall make use of this simple conception in the following section on prediction formulae.

Prediction Formulae. Except for simple cases such as for both parents‡ or for brethren§, I do not think many multiple regression formulae for kin have hitherto been worked out. In providing the present formulae I proceed from Equation (1), calculating the determinant and its minors for various groups of relatives, some direct, some collateral and some consisting partly of one and partly of the other. I assume that all individuals of the same grade have the same degree of resemblance, or if not that the mean of the correlations is taken. I shall assume females reduced to male equivalents by multiplying their deviations by the ratio of male to female

---

[*] Mean correlation of father and of mother with son.

[†] For use later I give the whole series of constants: $\alpha\beta = \cdot 556{,}6726$, $\alpha = \cdot 559{,}261$, $\theta = \cdot 703{,}834$, $\zeta = \cdot 716{,}078$, $\eta = \cdot 890{,}352$, $\gamma\eta = \cdot 105{,}020$, $\gamma = \cdot 117{,}953$.

[‡] First given, I think, in *R.S. Proc.* Vol. LVIII (1895), p. 240.

[§] See *Biometrika*, Vol. XVII, p. 136.

standard deviations. This is no real limitation, for it merely amounts to measuring the male and female deviations in terms of their respective variabilities.

Having obtained the regression equation, I shall suppose the deviation of each member of the group of relatives from the population type to be h times the standard deviation of his generation, and insert appropriate numerical values for the correlations. This is done in order to give the reader some numerical appreciation of the influence on a man's character of the possession of a certain number of relatives who deviate from the population type.

We are of course finding the average or type of *all* men who have such a group of relatives, and are supposing all the other relatives of the man to be mediocre, or of the population mean. This is, of course, unlikely to be the case. A man, for example, who has a father, an uncle and a grandfather of intellectual distinction, is likely to have other distinguished ancestors or collaterals, hence by putting all his other ancestors mediocre, we shall get a minimum prediction of his own probable grade of ability.

As a rule we only know those relatives who for any character differ markedly from type, and upon this knowledge we must base our prediction; the object of this investigation, however, is to show that if an individual has in his stirp a relatively few members deviating from mediocrity, it is quite easy for the prediction formula to give him a character value *in excess* of that of these noteworthy members. In other words, while inbreeding is the surest way to a continuously increasing deviation from type, still the union of members of two stirps, each of which has two or three members deviating in the same direction from type, will suffice to produce a progression of type, and this will continue with increasing speed so long as there is mating with like stocks.

In calculating the constants for the "Genometer" Plate (see p. 5) I adopted slightly different values for the parental and fraternal correlations, and consequently for the colligator, from those given in the tables below, which are based on the values for sons and brothers only in the case of stature in man. The changes are slight and of no importance, and it did not—as the whole scheme is merely illustrative—seem worth while to recompute for such small changes the genometer (which was originally prepared for Vol. III of my *Life of Francis Galton*).

We take here, as on p. 22, $r = \cdot 5110$, $r_1 = \cdot 5040$ and consequently $\beta = \cdot 995{,}372$, which leads to $\lambda = \cdot 695{,}758$. Further, $\epsilon = \cdot 2804$; α is determined from the second equation of (12) and we find for its value:

$$\alpha = \cdot 559{,}261,$$

whence, by (9), we obtain $\zeta = \cdot 716{,}078, \quad \theta = \cdot 703{,}834,$

and the individual pth grandparent has with the offspring a correlation given by

$$r_p = \cdot 5040 \times (\cdot 703{,}834)^{p-1}.$$

Thus we have

		[Values used on Genometer]
For Parents	$r_1 = \cdot 5040$	[·5062]
For Grandparents	$r_2 = \cdot 3547$	[·3563]
For Great Grandparents	$r_3 = \cdot 2497$	[·2519]
For Great Great Grandparents	$r_4 = \cdot 1757$	[·1781]
For Great Great Great Grandparents	$r_5 = \cdot 1237$	[·1259]
For Brothers	$r = \cdot 5110$	[·5337]
For Uncle and Nephew	$r_\lambda = \cdot 3555$	[·3686]
For First Cousins or Great Uncle and Nephew	$r_{\lambda^2} = \cdot 2474$	[·2579]
For a Man and his First Cousin once removed	$r_{\lambda^3} = \cdot 1721$	[·1804]
For Second Cousins	$r_{\lambda^4} = \cdot 1197$	[·1263]

It will be noted that owing to the near equality of parental (·5040) and fraternal (·5110) correlations, and the rates of reduction, $\theta = {\cdot}703{,}834$ and $\lambda = {\cdot}695{,}758$, we have very roughly $r_p = r_{\lambda^{p-1}}$, and accordingly it is difficult to determine whether a brother is closer in blood than a parent, a nephew than a grandparent, a cousin than a great-grandparent, and so on, the collaterals being at first closer than the direct ancestors.

Prediction Formulae.

$\tilde{x}_0$ = mean of all individuals having a relative or group of relatives of deviation $h_i\sigma_i$, where $h_i\sigma_i$ is the deviation from his own type of an individual i, whose generation has variability σ_i. $\bar{x}_0$ is the population mean of offspring generation.

(i) *A single parent, F or M.*

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}5040 h_F, \quad \text{or} \quad = \sigma_0 \times {\cdot}5040 h_M.$$

Hence for a single parent of deviation $h\sigma_F$ the progress from type is $\sigma_0 \times {\cdot}5040h$ [·5062]*.

(ii) *Both parents, F and M,* with assortative mating coefficient ϵ.

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \frac{r_1}{1+\epsilon}(h_F + h_M).$$

With values as above, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}393{,}627\,(h_F + h_M)$,

or, if $h_F = h_M = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}7873h$ [·7907].

(iii) *Both parents and one (say paternal) grandparent, F, M and G_p.* Here we need to determine the values of the minors of the determinant D, where

$$D = \begin{vmatrix} 1, & r_1, & r_1, & r_2 \\ r_1, & 1, & \epsilon, & r_1 \\ r_1, & \epsilon, & 1, & 0 \\ r_2, & r_1, & 0, & 1 \end{vmatrix}.$$

Here $D_{00} = 1 - \epsilon^2 - r_1^2$, $D_{01} = -\,r_1(1 - \epsilon - r_2)$,

$D_{02} = -\,r_1(1 - \epsilon - r_1^2 + \epsilon r_2)$, $D_{03} = -\,(1-\epsilon)\,((1+\epsilon)\,r_2 - r_1^2)$,

and we have

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0\left(\frac{r_1(1-\epsilon-r_2)}{1-\epsilon^2-r_1^2}h_F + \frac{r_1(1-\epsilon-r_1^2+\epsilon r_2)}{1-\epsilon^2-r_1^2}h_M + \frac{(1-\epsilon)\,((1+\epsilon)\,r_2 - r_1^2)}{1-\epsilon^2-r_1^2}h_{G_p}\right),$$

or turning into numbers,

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0\,({\cdot}274{,}369h_F + {\cdot}426{,}727h_M + {\cdot}215{,}807h_{G_p}),$$

and, if $h_F = h_M = h_{G_p} = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}9169h$ [·9215].

Thus there would still be a slight regression on the parental and grandparental value h.

(iv) *Both parents and two (say paternal) grandparents, F, M, G_p, $G_{p'}$.* Here

$$D = \begin{vmatrix} 1, & r_1, & r_1, & r_2, & r_2 \\ r_1, & 1, & \epsilon, & r_1, & r_1 \\ r_1, & \epsilon, & 1, & 0, & 0 \\ r_2, & r_1, & 0, & 1, & \epsilon \\ r_2, & r_1, & 0, & \epsilon, & 1 \end{vmatrix},$$

$D_{00} = (1-\epsilon)\,((1-\epsilon)\,(1+\epsilon)^2 - 2r_1^2)$, $D_{01} = -\,r_1(1-\epsilon)\,(1-\epsilon^2-2r_2)$,

$D_{02} = -\,r_1(1-\epsilon)\,(1-\epsilon^2-2r_1^2+2\epsilon r_2)$, $D_{03} = D_{04} = (1-\epsilon)^2\,(r_1^2 - r_2(1+\epsilon))$.

---

* Numbers in square brackets are the values used for the Genometer plate, where r and r_1 were arranged from several tables.

Hence we have for the regression equation the form

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r_1 (1 - \epsilon^2 - 2r_2)}{(1 - \epsilon)(1 + \epsilon)^2 - 2r_1^2} h_F + \frac{r_1 (1 - \epsilon^2 - 2r_1^2 + 2\epsilon r_2)}{(1 - \epsilon)(1 + \epsilon)^2 - 2r_1^2} h_M \right.$$
$$\left. + \frac{(1 - \epsilon)(r_2 (1 + \epsilon) - r_1^2)}{(1 - \epsilon)(1 + \epsilon)^2 - 2r_1^2} (h_{G_p} + h_{G_{p'}}) \right).$$

Or, inserting the above numerical values,

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 (\cdot 159{,}053 h_F + \cdot 459{,}400 h_M + \cdot 214{,}414 (h_{G_p} + h_{G_{p'}})).$$

Accordingly, if $h_F = h_M = h_{G_p} = h_{G_{p'}} = h$,

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times 1\cdot 0473 h \ [1\cdot 0516].$$

Thus the selection of both parents and the two grandparents on one side suffices to produce a slight advance on the selected value.

(v) *Both parents and one paternal and one maternal grandparent, F, M, G_p, G_m.* Here

$$D = \begin{vmatrix} 1, & r_1, & r_1, & r_2, & r_2 \\ r_1, & 1, & \epsilon, & r_1, & 0 \\ r_1, & \epsilon, & 1, & 0, & r_1 \\ r_2, & r_1, & 0, & 1, & 0 \\ r_2, & 0, & r_1, & 0, & 1 \end{vmatrix}$$

and $\Delta_{00} = (1 - r_1^2)^2 - \epsilon^2, \quad \Delta_{01} = \Delta_{02} = - r_1 (1 - r_2)(1 - r_1^2 - \epsilon),$

$$\Delta_{03} = \Delta_{04} = r_2 (1 - r_1^2 - \epsilon^2) - r_1^2 (1 - r_1^2 - \epsilon (1 - r_2)).$$

Hence

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r_1 (1 - r_2)(1 - r_1^2 - \epsilon)}{(1 - r_1^2)^2 - \epsilon^2} (h_F + h_M) + \frac{r_2 (1 - r_1^2 - \epsilon^2) - r_1^2 (1 - r_1^2 - \epsilon (1 - r_2))}{(1 - r_1^2)^2 - \epsilon^2} (G_p + G_m) \right).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 (\cdot 316{,}871 (h_F + h_M) + \cdot 194{,}997 (G_p + G_m)),$

and, if $h_F = h_M = G_p = G_m$, we have

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times 1\cdot 0237 h \ [1\cdot 0268].$$

We note again how few relatives deviating from normality suffice to carry a man up to and beyond the grade of these ancestors. An interesting point is worth noting, an individual is in a slightly better position if, with two grandparents of distinction, they are on the same parental side.

(vi) *All four grandparents, G_p, $G_{p'}$, G_m, $G_{m'}$.* We have for D,

$$D = \begin{vmatrix} 1, & r_2, & r_2, & r_2, & r_2 \\ r_2, & 1, & \epsilon, & 0, & 0 \\ r_2, & \epsilon, & 1, & 0, & 0 \\ r_2, & 0, & 0, & 1, & \epsilon \\ r_2, & 0, & 0, & \epsilon, & 1 \end{vmatrix}.$$

Here $\Delta_{00} = (1 - \epsilon^2)^2, \quad \Delta_{01} = \Delta_{02} = \Delta_{03} = \Delta_{04} = r_2 (1 - \epsilon)(1 - \epsilon^2),$

and accordingly $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \dfrac{r_2}{1 + \epsilon} (h_P + h_{P'} + h_M + h_{M'}),$

or numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 277{,}023 (h_P + h_{P'} + h_M + h_{M'}),$

and, if $h_P = h_{P'} = h_M = h_{M'} = h$,

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times 1\cdot 1081 h \ [1\cdot 1132],$$

or four grandparents of excess from type will carry a grandson on the average about 11 % beyond their average excess.

(vii) *Three grandparents, G_p, $G_{p'}$ and G_m.* We obtain for D,

$$D = \begin{vmatrix} 1, & r_2, & r_2, & r_2 \\ r_2, & 1, & \epsilon, & 0 \\ r_2, & \epsilon, & 1, & 0 \\ r_2, & 0, & 0, & 1 \end{vmatrix},$$

$$D_{00} = 1 - \epsilon^2, \quad D_{01} = D_{02} = -r_2(1-\epsilon), \quad D_{03} = -r_2(1-\epsilon^2).$$

Hence

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0\left(\frac{r_2}{1+\epsilon}(h_p + h_{p'}) + r_2 h_m\right) = \sigma_0\,(\cdot 277{,}023\,(h_p + h_{p'}) + \cdot 354{,}700 h_m).$$

If $h_p = h_{p'} = h_m = h$, $\qquad \tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 9087h\ [\cdot 9128]$.

(viii) *Both parents and all four grandparents, F, M, G_p, $G_{p'}$, G_m, $G_{m'}$.* We have

$$D = \begin{vmatrix} 1, & r_1, & r_1, & r_2, & r_2, & r_2, & r_2 \\ r_1, & 1, & \epsilon, & r_1, & r_1, & 0, & 0 \\ r_1, & \epsilon, & 1, & 0, & 0, & r_1, & r_1 \\ r_2, & r_1, & 0, & 1, & \epsilon, & 0, & 0 \\ r_2, & r_1, & 0, & \epsilon, & 1, & 0, & 0 \\ r_2, & 0, & r_1, & 0, & 0, & 1, & \epsilon \\ r_2, & 0, & r_1, & 0, & 0, & \epsilon, & 1 \end{vmatrix}$$

and we find

$$\Delta_{00} = (1-\epsilon)^2\,((1+\epsilon-2r_1{}^2)^2 - \epsilon^2(1+\epsilon)^2),$$

$$\Delta_{01} = -(1-\epsilon)^2\,(r_1(1+\epsilon) - 2r_1r_2)(1-\epsilon^2-2r_1{}^2) = \Delta_{02},$$

$$\Delta_{03} = -(1-\epsilon)^2\,[\epsilon r_1{}^2(1-\epsilon^2) - 2\epsilon r_1{}^2 r_2 + (r_2 - r_1{}^2)((1-\epsilon^2)(1+\epsilon) - 2r_1{}^2)]$$
$$= \Delta_{04} = \Delta_{05} = \Delta_{06}.$$

Accordingly

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0\left(\frac{r_1(1+\epsilon) - 2r_1r_2}{(1+\epsilon)^2 - 2r_1{}^2}(h_F + h_M) + \frac{r_2(1+\epsilon) - r_1{}^2}{(1+\epsilon)^2 - 2r_1{}^2}(h_p + h_{p'} + h_m + h_{m'})\right).$$

Numerically

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0\,(\cdot 254{,}363\,(h_F + h_M) + \cdot 176{,}898\,(h_p + h_{p'} + h_m + h_{m'})),$$

and, if $h_F = h_M = h_p = h_{p'} = h_m = h_{m'} = h$,

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times 1\cdot 2163h\ [1\cdot 2199].$$

Thus if an individual have parents and grandparents only deviating h from type, and no further selection of ancestry, then he will on the average progress 22 % beyond these selected relatives for this character. The advantage of inbreeding thus becomes manifest.

(ix) *A maternal and a paternal grandparent, G_m and G_p.* We have

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 r_2\,(h_m + h_p),$$

and numerically $\qquad \tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 3547\,(h_m + h_p)$.

If $h_m = h_p = h$, $\qquad \tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 7094h\ [\cdot 7126]$.

(x) *Two grandparents, husband and wife,* G_p, $G_{p'}$.

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \frac{r_2}{1 + \epsilon} (h_p + h_{p'}).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}277{,}023\,(h_p + h_{p'})$,

or, if $h_p = h_{p'} = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}5540h\ [{\cdot}5565]$.

(xi) *Two great-grandparents, husband and wife,* G_{p^2}, $G_{p'^2}$.

$$\tilde{x}_0 - \bar{x} = \sigma_0 \frac{r_2}{1 + \epsilon} (h_{p^2} + h_{p'^2}).$$

Numerically $\tilde{x}_0 - \bar{x} = \sigma_0 \times {\cdot}195{,}017\,(h_{p^2} + h_{p'^2})$,

or, if $h_{p^2} = h_{p'^2} = h$, $\tilde{x}_0 - \bar{x} = \sigma_0 \times {\cdot}3900h\ [{\cdot}3935]$.

(xii) *Father and paternal grandparents,* F, G_p, $G_{p'}$. The determinant is here

$$D = \begin{vmatrix} 1, & r_1, & r_2, & r_2 \\ r_1, & 1, & r_1, & r_1 \\ r_2, & r_1, & 1, & \epsilon \\ r_2, & r_1, & \epsilon, & 1 \end{vmatrix},$$

$$D_{00} = (1 - \epsilon)(1 + \epsilon - 2r_1^2), \quad D_{01} = -r_1(1 - \epsilon)(1 + \epsilon - 2r_2),$$

$$D_{02} = D_{03} = -(1 - \epsilon)(r_2 - r_1^2),$$

and accordingly

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r_1(1 + \epsilon - 2r_2)}{1 + \epsilon - 2r_1^2} h_F + \frac{r_2 - r_1^2}{1 + \epsilon - 2r_1^2} (h_p + h_{p'}) \right).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 ({\cdot}372{,}600h_F + {\cdot}132{,}429\,(h_p + h_{p'}))$,

or, if $h_F = h_p = h_{p'} = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}6375h\ [{\cdot}6351]$.

(xiii) *Father, paternal grandfather and the latter's father,* F, G_p, G_{p^2}. Here

$$D = \begin{vmatrix} 1, & r_1, & r_2, & r_3 \\ r_1, & 1, & r_1, & r_2 \\ r_2, & r_1, & 1, & r_1 \\ r_3, & r_2, & r_1, & 1 \end{vmatrix},$$

$$\Delta_{00} = (1 - r_2)(1 + r_2 - 2r_1^2), \quad \Delta_{01} = -((1 - r_2)(r_1 - r_2 r_3) - (r_1 - r_3)(r_1^2 - r_2^2)),$$

$$\Delta_{02} = (1 - r_2)(r_1^2 - r_2^2 - r_2 + r_1 r_3), \quad \Delta_{03} = (1 - r_2)^2 r_1 + (1 - r_1^2)(r_3 - r_1),$$

and accordingly

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{(1 - r_2)(r_1 - r_2 r_3) - (r_1 - r_3)(r_1^2 - r_2^2)}{(1 - r_2)(1 + r_2 - 2r_1^2)} h_F \right.$$
$$\left. + \frac{(1 - r_2)(r_1^2 - r_2^2 - r_2 + r_1 r_3)}{(1 - r_2)(1 + r_2 - 2r_1^2)} h_p + \frac{r_1(1 - r_2)^2 - (r_1 - r_3)(1 - r_1^2)}{(1 - r_2)(1 + r_2 - 2r_1^2)} h_{p^2} \right).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 ({\cdot}430{,}994h_F + {\cdot}118{,}875h_p + {\cdot}036{,}913h_{p^2})$,

or, if $h_F = h_p = h_{p^2} = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times {\cdot}5868h\ [{\cdot}5889]$.

This result, compared, say, with (ix) or (xii), indicates how far more profitable it is that noteworthiness should not be confined to one line of ascent, i.e. if a man has three or four distinguished ancestors it is far better that they should not all bear his own surname*. Of course this is an obvious corollary from the general principle that prediction should be determined from variables $x_1, x_2 \ldots x_n$, highly correlated with x_0, but not with each other. But the eugenic principle which follows, namely, that the individual with distinguished ancestors of differing birth surnames is more likely to be himself distinguished is not always borne in mind.

I now turn to prediction from collaterals, and here it might seem that brothers are the most important. In a certain sense they are, a noteworthy uncle is of less value than a noteworthy brother, but two noteworthy uncles, if of maternal and paternal lines, are of sensibly more value than two noteworthy brothers.

(xiv) *Any number n of brothers, B, B', B''....* The general formula for the prediction of type in an individual x_0 from $x_1, x_2 \ldots x_n$, when all the x's are equally correlated, is well known. In our particular case it takes the form, r being the correlation of brothers,

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \frac{r}{1 + (n-1)\,r} \overset{p=n}{\underset{p=1}{S}} (h_{b_p}) = \sigma_0 \frac{nr}{1 + (n-1)\,r} \text{ (mean } h_b \text{ for brothers of } x_0).$$

Hence, numerically, since $r = \cdot 5110$:

For a single brother ... ... $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 5110 h_b$
For two brothers ... ... $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 338{,}187\,(h_b + h_{b'})$
,, ,, if $h_b = h_{b'} = h$ $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 6764 h$
For three brothers ,, ,, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 7582 \bar{h}_b$
For four brothers ,, ,, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 8069 \bar{h}_b$
For five brothers ,, ,, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 8394 \bar{h}_b$ and so on.

An indefinitely large number of brothers could, however, never carry on the average of the brother x_0 above the mean of the predicating brothers $x_1, x_2 \ldots x_n$. The possibility of progression depends on ancestry.

(xv) *Any number of uncles, u_p, $u_{p'}$, $u_{p''}$ ....* Provided the uncles are all paternal or all maternal the problem is precisely like that of brothers. We have only to replace the r in the numerator by $r\lambda$, where λ is the colligator, or

$$\tilde{x}_0 - \bar{x}_0 = \frac{\sigma_0 r\lambda}{1 + (n-1)\,r} \overset{p=n}{\underset{p=1}{S}} (h_{u_p}) = \frac{\sigma_0 n r\lambda}{1 + (n-1)\,r} \bar{h}_u .$$

Hence, numerically, since $r\lambda = \cdot 695{,}758$:

For a single paternal uncle ... ... $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 3555 \bar{h}_u$
For two paternal uncles ... ... $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 4706 \bar{h}_u$
For three paternal uncles ... ... $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 5275 \bar{h}_u$
For four paternal uncles ... ... $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 5614 \bar{h}_u$ and so on.

Clearly an increasing number of noteworthy brothers of the father does not mean rapid rise in type, four uncles on one side being far less valuable than two grandparents on different sides.

(xvi) *A maternal uncle and a paternal uncle, u_p and u_m.* The formula is now

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times r\lambda\,(h_{u_p} + h_{u_m}).$$

---

* An exception, if slight, to this rule appears in (iv) and (v). I have not been able to find any error in the algebra or the arithmetic.

Numerically $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 3555\,(h_{u_p} + h_{u_m}),$$

and, if $h_{u_p} = h_{u_m} = h$, $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 7110 h,$$

or two uncles, maternal and paternal, are about 27 % better than four uncles all paternal.

(xvii) *Two maternal and two paternal uncles, u_p, $u_{p'}$, u_m, $u_{m'}$.* Here we have to deal with a 5 × 5 determinant, namely,

$$D = \begin{vmatrix} 1, & r_\lambda, & r_\lambda, & r_\lambda, & r_\lambda \\ r_\lambda, & 1, & r, & 0, & 0 \\ r_\lambda, & r, & 1, & 0, & 0 \\ r_\lambda, & 0, & 0, & 1, & r \\ r_\lambda, & 0, & 0, & r, & 1 \end{vmatrix},$$

$$D_{00} = (1 - r^2)^2, \quad D_{01} = D_{02} = D_{03} = D_{04} = r_\lambda (1 - r)(1 - r^2).$$

Hence, since $r_\lambda = r\lambda$, $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r\lambda}{1 + r} (h_{u_p} + h_{u_{p'}} + h_{u_m} + h_{u_{m'}}) \right).$$

Numerically $$\tilde{x} - \bar{x}_0 = \sigma_0 \times \cdot 235{,}275\,(h_{u_p} + h_{u_{p'}} + h_{u_m} + h_{u_{m'}}),$$

and, if $h_{u_p} = h_{u_{p'}} = h_{u_m} = h_{u_{m'}} = h$,
$$\tilde{x} - \bar{x}_0 = \sigma_0 \times \cdot 9411 h.$$

A comparison with the result for four paternal uncles shows that to have two uncles, exceptional in both lines, is about 68 % better than having four exceptional uncles in one line of descent.

We will emphasise this in the following case.

(xviii) *Two paternal and one maternal uncle, u_p, $u_{p'}$ and u_m.* Here the regression equation is

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r\lambda}{1 + r} (h_{u_p} + h_{u_{p'}}) + r\lambda h_{u_m} \right).$$

Numerically $$\tilde{x}_0 - \bar{x}_0 = \sigma_0\,(\cdot 235{,}275\,(h_{u_p} + h_{u_{p'}}) + \cdot 3555 h_{u_m}).$$

If $h_{u_p} = h_{u_{p'}} = h_{u_m} = h$, then $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 8251 h.$$

Or, in the case of three noteworthy uncles, if one comes from the maternal side the type of nephew will be 56 % superior to the type when all three uncles are on the paternal side.

I now pass to prediction from cousins.

(xix) *One cousin, C.*
$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 r \lambda^2 h_c.$$

Or, numerically, $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 2474 h.$$

(xx) *Two cousins, who are brothers, C, C′.*
$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \frac{r\lambda^2}{1 + r} (h_c + h_{c'}).$$

Numerically $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 163{,}733\,(h_c + h_{c'}).$$

If $h_c = h_{c'} = h$, $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 3375 h.$$

(xxi) *Two cousins, one on either parental side,* C_p, C_m.

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 r\lambda^2 \left(h_{c_p} + h_{c_m}\right).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 2474 \left(h_{c_p} + h_{c_m}\right).$

If $h_{c_p} = h_{c_m} = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 4948h.$

(xxii) *Four cousins, not brothers, two on either side,* C_p, $C_{p''}$, C_m, $C_{m''}$. Each pair will be cousins of each other as well as of the subject, but uncorrelated with the members of the other pair.

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r\lambda^2}{1 + r\lambda^2} \left(h_{c_p} + h_{c_{p''}} + h_{c_m} + h_{c_{m''}}\right)\right).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 198{,}364 \left(h_{c_p} + h_{c_{p''}} + h_{c_m} + h_{c_{m''}}\right),$

and, if $h_{c_p} = h_{c_{p''}} = h_{c_m} = h_{c_{m''}} = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 7935h.$

Comparing with (ii) we see that four noteworthy cousins, two from each side, are of slightly more value than two noteworthy parents.

It is usually supposed that very little of the hereditary characters of an individual can be learnt from a man's cousins. This is not the case, if the cousins be judiciously chosen so as not to be brothers and taken from maternal or paternal stock only.

Lastly, I will illustrate in two or three cases combined prediction from ancestors and collaterals.

(xxiii) *A parent and a brother, F, B.* Here

$$D = \begin{vmatrix} 1, & r_1, & r \\ r_1, & 1, & r_1 \\ r, & r_1, & 1 \end{vmatrix},$$

$$D_{00} = 1 - r_1^2, \quad D_{01} = -r_1(1 - r), \quad D_{02} = -(r - r_1^2).$$

Hence $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r_1(1 - r)}{1 - r_1^2} h_p + \frac{r - r_1^2}{1 - r_1^2} h_b\right),$$

and, numerically, $= \sigma_0 (\cdot 330{,}377 h_p + \cdot 344{,}490 h_b).$

If $h_p = h_b = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 6749h.$

(xxiv) *Two parents and one brother, F, M, B.* Here

$$D = \begin{vmatrix} 1, & r_1, & r_1, & r \\ r_1, & 1, & \epsilon, & r_1 \\ r_1, & \epsilon, & 1, & r_1 \\ r, & r_1, & r_1, & 1 \end{vmatrix},$$

$$D_{00} = (1 + \epsilon - 2r_1^2)(1 - \epsilon), \quad D_{01} = -r_1(1 - r)(1 - \epsilon) = D_{02},$$

and $$D_{03} = -(1 - \epsilon)((1 + \epsilon) r - 2r_1^2).$$

Thus $$\tilde{x}_0 - \bar{x}_0 = \sigma_0 \left(\frac{r_1(1 - r)}{1 + \epsilon - 2r_1^2} (h_F + h_M) + \frac{(1 + \epsilon) r - 2r_1^2}{1 + \epsilon - 2r_1^2} h_B\right).$$

Numerically $\tilde{x}_0 - \bar{x}_0 = \sigma_0 (\cdot 319{,}091 (h_F + h_M) + \cdot 189{,}356 h_B).$

If $h_F = h_M = h_B = h$, $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 8275h.$

(xxv) *Two parents and two brothers, F, M, B and B′.* The object of this particular group is to indicate how very little is gained by adding an additional brother, since he is highly correlated with both the parents and the other brother from whom prediction is to be made. Here

$$D = \begin{vmatrix} 1, & r_1, & r_1, & r, & r \\ r_1, & 1, & \epsilon, & r_1, & r_1 \\ r_1, & \epsilon, & 1, & r_1, & r_1 \\ r, & r_1, & r_1, & 1, & r \\ r, & r_1, & r_1, & r, & 1 \end{vmatrix}.$$

Hence

$$D_{00} = (1 - \epsilon)(1 - r)((1 + \epsilon)(1 + r) - 4r_1^2), \quad D_{01} = D_{02} = -r_1(1 - \epsilon)(1 - r)^2,$$

$$D_{03} = D_{04} = -(1 - \epsilon)(1 - r)((1 + \epsilon)r - 2r_1^2).$$

Hence

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0\left(\frac{r_1(1 - r)}{(1 + \epsilon)(1 + r) - 4r_1^2}(h_F + h_M) + \frac{(1 + \epsilon)r - 2r_1^2}{(1 + \epsilon)(1 + r) - 4r_1^2}(h_B + h_{B'})\right).$$

Numerically

$$\tilde{x}_0 - \bar{x}_0 = \sigma_0(\cdot 268{,}289(h_F + h_M) + \cdot 159{,}208(h_B + h_{B'})).$$

If $h_F = h_M = h_B = h_{B'} = h$, then $\tilde{x}_0 - \bar{x}_0 = \sigma_0 \times \cdot 8550h$,

which is only about a $3\frac{1}{3}$ % higher value than if one brother be omitted as in (xxiv).

We are now in a position to draw up a table for stature in man indicating how, when we increase the number and raise the degree of affinity of the group of relatives who have excess of deviation from type in the stirp of individuals, we gradually raise their average deviation from type until it ultimately equals and then excels the deviation of the group of relatives themselves. The importance of this result for eugenics is great; it indicates that while inbreeding will most rapidly increase deviation (good or bad) from type, still judicious assortative mating—that is, mating which pays attention on both sides to the number of atypical members of the two stirps—is capable not only of maintaining past capacity for achievement, but actually of carrying the new combined stirps beyond their past deviation from mediocrity.

One point in particular the present results emphasise. Better results will accrue from the intermarriage of two stocks, each of which has *n* noteworthy members, than from the marriage with a mediocre stock of a stock which has 2*n* noteworthy members.

I should like finally to emphasise in this Appendix a point to which I have referred often in my lectures, but which is too often overlooked. When a Mendelian asserts that a particular individual is homozygous for a particular character, on what is his knowledge actually based? He is clearly applying the word to the identity of the factor for this character in all that individual's germ cells. But those germ cells are not *a priori* open to his inspection, even if he could, from an inspection, ascertain this identity. He is therefore basing his assertion on something else, which he can observe, that is to say on the observation of somatic characters. We naturally ask, *what* somatic characters? Clearly the individual's own somatic characters would be a quite inadequate basis for any assertion as to the absolute identity of a factor in the germ cells. Actually the statement is based upon either (i) a knowledge of the somatic characters of a portion of the ancestry,—the Mendelian can never know the somatic characters of the whole ancestry,—or (ii) a knowledge of what some similar somatic individuals have produced in the past, again not a knowledge of *all* similar individuals. In either case what his assertion amounts to is a probability based not on gametic but on somatic characters of kin. The probability is similar in character to that which the Ancestral Law *measures.* The geneticist who is not acquainted with the idea that all assertions, even that of the sun rising to-morrow, are ultimately based on probability may say that he has mated A with A' and they are homozygous because their offspring give only A's, A being a certain character, for which the homozygosity is asserted. To this the answer is clear. In the first place, he is making an assertion based not on the gametic characters of A and A' but on the somatic characters of an enlarged pedigree, ancestors and collaterals of A's. He can never test the whole germinal possibilities of either A and A'. What if 1 % or even 0·1 % of the germinal cells of A are unlike their fellows*? How long may it be before chance brings them into evidence, and if it does will they be conveniently dismissed as "rogues"? In the second and perhaps more important place, what does the geneticist mean by a somatic character A? Until he has placed

Table of the numerical factors to be multiplied by $h\sigma_0$, which mark the average progression (or regression) from mediocrity of individuals having the relatives recorded below deviating from type by $h\sigma$.

Index to Case	Group of Relatives	Factor
(viii)	Two Parents and Four Grandparents	1·2163
(vi)	All four Grandparents	1·1081
(iv)	Two Parents and the two Paternal Grandparents	1·0473
(v)	Two Parents, one Paternal and one Maternal Grandparent	1·0237
(xvii)	Two Paternal and two Maternal Uncles	·9411
(iii)	Two Parents and one Paternal Grandparent	·9169
(vii)	Three Grandparents	·9087
(xxv)	Two Parents and two Brothers	·8550
(xiv)	Five Brothers	·8394
(xxiv)	Two Parents and one Brother	·8275
(xviii)	Two Paternal and one Maternal Uncle	·8251
(xxii)	Four Cousins, not brothers, two on either side	·7935
(ii)	Two Parents	·7873
(xiv)	Three Brothers	·7582
(xvi)	A Maternal and a Paternal Uncle	·7110
(ix)	A Maternal and a Paternal Grandparent	·7094
(xiv)	Two Brothers	·6764
(xxiii)	A Parent and a Brother	·6749
(xii)	Father and Paternal Grandparents	·6375
(xiii)	Father, Paternal Grandfather and latter's Father	·5868
(xv)	Four Paternal Uncles	·5614
(x)	Both Paternal Grandparents	·5540
(xiv)	A Brother	·5110
(i)	A Parent	·5040
(xxi)	Two Cousins, one on either parental side	·4948
(xv)	Two Paternal Uncles	·4706
(p. 28 ftn.)	Two Half-Brothers	·4024
(xi)	Two Great Grandparents, Husband and Wife	·3900
(xv)	An Uncle	·3555
(p. 23)	A Grandparent	·3547
(xx)	Two Cousins, Brethren	·3375
(p. 28)	A Half-Brother	·2519
(xix)	One Cousin	·2474
(p. 23)	A Great Grandparent	·2497
(p. 23)	A Great Great Grandparent	·1757
(p. 29)	First Cousin once removed	·1721
(p. 23)	A Great Great Great Grandparent	·1237
(p. 29)	A Second Cousin	·1197

---

* You may cross an albino Pekinese with a pure bred (!) black Pomeranian and deduce the conclusion that the resulting "Pompek" will invariably be a black dog with more or less white shirt front, and feel then capable of *asserting* that black coat dominates white; but ultimately a chocolate dog will appear to demonstrate how idle it is to replace probability by assertion.

some measure upon it, it is impossible to say whether all A's are the same, or without further breeding whether the gametic constitution of all individuals bearing the so-called A-somatic character are alike. Here arises the importance of measuring and classifying within the A-character, and ultimately, as referred to in the lecture now printed, the possibility of differentiating within the "pure line." As the biometricians suggested at the end of the 19th century loose categories: "smooth and wrinkled," "tall and short," "hairy and smooth," "albino and normal," et hoc totum genus—omnis commeatus totius-que belli apparatus—would be the great stumbling-blocks in the ultimate progress of the Mendelian. Certainly if "albinism" be not continuous, it connotes in dogs several, perhaps many factors; in hackneys the various shades of "chestnut" coat are inherited, while in mice "piebaldism" is a graded character and as such inherited.