THE CAUSES OF HUMAN VARIABILITY.[1]

By R. A. FISHER.

THE great service which the modern development of statistics has rendered to eugenics is that it supplies a definite method of measuring and analysing variability. Only so can the true causes of variability be ascertained and the factors which are of no effect, the false claimants to importance in this regard, be excluded from consideration.

The practical method of measuring the variability of a population is extremely simple, although explanations of it are apt to appear complicated at a first reading. If a measurement of stature, for example, is made upon a sufficiently large sample—1,000 to 10,000, let us say, of a population—the measurements are usually found to be grouped symmetrically about a mean value, the average stature of the sample. The deviations from this average follow very closely in their distribution what is known as the law of errors, that is to say the frequencies with which deviations of different magnitudes occur are related in the same way as the frequencies of errors of observation. Consequently the amount of variability may be measured, as errors of observation are habitually measured, by the mean of the squares of the deviations of different individuals from the mean, this mean square deviation being strictly comparable to the mean square error used in astronomy and in all physical science.

Thus from 1,052[1] measurements of stature, Karl Pearson and Alice Lee found a mean stature 62·48[1] inches, and a mean square deviation 5·71[2] square inches. This mean square deviation I term the variance, and use it as the measure of variability, by reason of this important property, namely, that two independent causes of variability acting together produce a variance which is the sum of the variances produced by either separately. The square root of the variance has long been known as the standard deviation. The variance, then, is a measurable feature of the

[1] Trans. Roy. Soc. Edin. Vol. LII., Part II. (No. 13), pp. 399-434. The Correlation between Relations on the Supposition of Mendelian Inheritance.

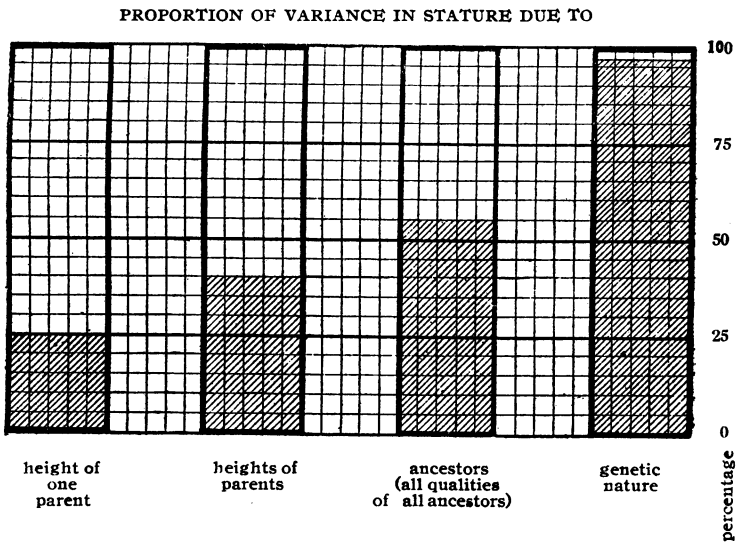
[2] These measurements are those of the mothers, of whom the number measured is given.

population, and indicates the extent to which its members differ one from another; applied to an individual it is meaningless, although each individual measurement is taken into account in calculating it. In the same way it should be clearly understood what we mean by a *cause of variability*. If we say, "This boy has grown tall because he has been well fed," we are not merely tracing out cause and effect in an individual instance; we are suggesting that he might quite probably have been worse fed, and that in this case he would have been shorter. We are, in fact, suggesting that existing differences of nutrition can account for differences of stature comparable to the standard deviation of stature. Now this is just what is meant when we speak of nutrition as a cause of variability; we thereby mean that in a population absolutely uniform in regard to other causes, such as breeding and exercise, existing differences of nutrition would produce a certain variability—in fact, that *a certain percentage* of the variance must be ascribed to nutrition.

This point of view throws a flood of light upon the meaning to be attached to the results of biometrical research. The statement that the correlation, in stature, between father and son is r may be interpreted in several ways. For our present purpose it means that r^2 represents the fraction of the total variance of the sons, which is due to the fact that their fathers differ in height. A selected population of sons of fathers of identical height would have its variance reduced by this fraction. Since r is about ·51 we may say that 26 per cent. of the variance is due to the stature of the father; the remaining 74 per cent. must have some other cause. Now this very value for the parental correlation has been freely used to prove that the influence of heredity is all-important. It has been contrasted with correlations such as ·05 obtained from particular environmental conditions. These latter evidently make no sensible contribution to the total variance, but it is just as evident that inheritance, although making an extensive contribution, has not been shown to furnish so much that there may not be other factors, equally or more important. An advocate of the importance of environment might easily point to a dozen causes to which height or shortness is commonly ascribed, such as regular athletic exercise, or accidental illness in childhood, and it would

be difficult to prove without a specially designed investigation for each alleged cause that these do not contribute important proportions of the total. The task of ascertaining the importance of environment in this way is an endless one, since always new environmental causes could be suggested, each more difficult than the last to define, measure and investigate.

On the other hand we have evidently not done justice to the importance of inborn character by considering merely the contribution of a single parent. We must obviously also take the mother into consideration. The maternal correlation is almost exactly the same as the paternal, and if father and mother were uncorrelated



we should have simply to double the amount of variance already accounted for. There is, however, in stature, quite a large marital correlation; so that the influence of the mother has already to some extent been allowed for in the paternal correlation. Taking the paternal and maternal correlations both equal to ·5066, and the marital correlation equal to ·2804, calculation shows that *the heights of the parents alone* account for 40·10 per cent. of the observed variance.

It is clear, however, that in taking the parental heights merely, we are not getting at the whole influence of parentage. It has long been known that the grandparents make an independent

contribution, and since this influence is transmitted by the parents it shows that the whole parental influence is not revealed by their stature alone. If the correlations for successive groups of ancestors were known, it would be possible to increase the 40·10 per cent. of the parents by taking account of the stature of the whole ancestry. Even this would not be sufficient, for characteristics of the organism not necessarily associated with stature in the ancestry might easily exert an influence upon it in the son. Stature is but one feature, but the whole organisation is derived from the parents. It is reasonable to suppose that children of different parents, chosen to have the same stature, would resemble each other less than children of the same actual parents, in point of stature as well as in other ways.

The only direct way of ascertaining the total effect of ancestry is, in fact, by measuring the amount of divergence between brothers and sisters of the same fraternity. This is what is done in ascertaining the fraternal correlation, and it may be shown that the fraternal correlation itself measures the fraction of the variance for which parentage, including ancestry, is responsible. The fraternal correlation, taken from the same data as before, is ·5433 ; hence, although only 40·10 per cent. could be ascribed to the stature of parents, *the total effect of parentage* is responsible for 54·33 per cent.

This result precipitates the enquiry, " If parentage only causes 54 per cent. of the variance, to what is the remaining 46 per cent due? Is there any other important cause of the observed differences, besides heredity? Is it possible that environment, acting in some subtle ways undetected by statisticians, can have so large an influence?" To these queries, in the case of stature, the facts appear to give a decided negative. Brothers show the same decided differences whether the experience of their early life is closely similar or very different. The argument in favour of the preponderant importance of heredity is enforced by Galton's observations of identical twins ; for he found that in certain cases twins resemble one another very decidedly more than do ordinary brothers and sisters, and he suggested that these twins, developing from the same zygote, partake of the same innate constitution, whereas brothers and sisters, including ordinary twins, do not possess the same hereditary qualities.

The fact is that qualities which are hereditary in the strictest sense, need not necessarily be inherited. In Mendelism, the cross breed, or heterozygote, hands on one quality to half his offspring and another alternative quality to the other half. Children of the same parentage may therefore differ among themselves by the *segregation of hereditary qualities.* It may be that features such as stature, in the distribution of which no discontinuity is apparent, may in reality be determined by a large number of Mendelian factors. These factors no doubt affect other features as well, and this would account for the high correlations which are known to exist between different measurements of the same individual; but whatever other effects they may have, each may be supposed to exert a small and definite influence upon stature. This theory has been previously discussed by Pearson[1] and other writers, and it appeared to be believed that the facts of biometry were not susceptible of such an interpretation. The theory, however, which I had the advantage of discussing on several occasions with Major Darwin, appeared to me so well worth a thorough investigation that I decided to make a detailed mathematical study of the biometrical properties of such a population. The methods and analysis have been published in a recent paper,[2] and the results are far more simple and definite than I had dared to hope.

Pearson had found that the assumption of Mendelian heredity led him to the low value, $\frac{3}{5}$ for the parental correlation and various values not exceeding $\frac{1}{12}$ for the fraternal correlation. These were obviously far below the true figures, but it was subsequently suggested by Yule (Conference on Genetics, 1906) that both parental and fraternal correlations could be raised to $\frac{1}{2}$ by abandoning the assumption of perfect dominance; with this view Pearson subsequently agreed. When I came to investigate the point I took care, therefore, to make no assumptions about dominance. Yule had suggested that dominance had an effect analogous to environment, or to arbitrary errors introduced

[1] Mathematical Contributions to the Theory of Evolution, XII. On a generalised theory of alternative inheritance with special reference to Mendel's laws. Phil. Trans., 1903.

[2] The Correlation between Relations on the Supposition of Mendelian Inheritance. Trans. Royal Soc., Edin., Vol. LII., Part II., pp. 399-433.

into the readings, and I soon found that on the most general assumptions I could make the whole effect of dominance could be represented by a single fraction, by which the *parental correlation* was reduced. This confirmed Yule's suggestion, but I was astonished to find that the *fraternal correlation* was reduced only to half the same extent. In this respect, then, the effect of dominance differs from that of arbitrary errors; for the latter reduce all correlations alike. By this means the two reduction factors may be separated, and we may estimate how much of the reduction of the parental correlation is due to dominance and how much to "environment," using this term to include all non-genetic causes of variability.

Before this could be done it was necessary to face the greatest analytical difficulty of the problem, viz., to make allowance for the correlation of father and mother. This correlation has the most profound effects upon the distribution of factors in the population, for, by virtue of it, similar phases of different factors become associated; the taller of two alternatives occurs in the same individual with the taller of another pair more frequently than random association would allow. Consequently the variance is increased, the extremes of high and low stature becoming more common. With stature I find the variance increased from 5·611 sq. inches to 6·760 sq. inches,[1] so that 17 per cent. of the total variance is due to the mating of like with like.

Another effect is to increase the ratio between the correlations of successive ancestors. The correlations diminish, in accordance with the "Law of Ancestral Heredity" in geometrical progression. Without homogamy the common ratio would be $\frac{1}{2}$, but for eye colour it is known to be as high as ·62, and for stature its value, calculated from the marital correlation, is ·6205.

The dominance ratio, which, in the absence of homogamy, is the fraction by which the parental correlation is reduced by the effect of dominance, does not appear to be less than would accord with complete dominance. For stature I find the value to be ·276, whereas the value upon Pearson's assumptions of complete dominance and equal frequency of the alternative types of gamete should be ·333; the ratio may be reduced not only by incomplete

[1] These are mean values for fathers, mothers, sons and daughters.

dominance, but by one type of gamete being more or less frequent than its alternative. If this inequality is not weighted in favour of rare recessives, then the dominance ratio is diminished. I see no reason to postulate incomplete dominance.

The parental and fraternal correlations are raised to high values about $\frac{1}{2}$, not by the abandonment of dominance, which would make them equal, but by homogamy, which raises their values while maintaining the difference between them.

The total proportion of variance which is due to innate and heritable causes may be calculated from the difference between the fraternal and the parental calculations; this is a small difference, and unfortunately, in our present data, errors of random sampling may modify it by a considerable proportion. The estimate of the possible effect of non-heritable factors, which is small, is consequently smaller than the probable error of this determination. The factor representing the proportionate effect of innate causes is actually greater than unity, but exceeds it by an amount less than its standard error; the actual figure is 1·01, with a standard error ·08; the value is therefore very unlikely to be less than ·93, and we may take ·95 as a low estimate. This conclusion is strengthened by the fact that the mean of three determinations of this constant for three different features is ·982.

There is, then, in this analysis of variance no indication of any other than innate and heritable factors at work, and a strong probability that whatever non-heritable factors are at work, including errors of measurement, do not contribute more than 5 per cent. and perhaps much less, to the total variance. Any such residue could be determined by a somewhat more accurate numerical knowledge of the relationship correlations, and in particular of the fraternal correlation. Fortunately this is of all the correlation coefficients the one for which it is easiest to gather data, and it is, in my opinion, the one with the simplest and most important meaning.

The conclusions of my investigation may be summed up as follows:—

(1) The facts of Biometry do not contradict, but in many ways positively support the theory of cumulative Mendelian factors.

(2) If this theory is correct a sufficient knowledge of the correlation coefficients for any one feature, between different pairs of relatives, would enable us to analyse completely and estimate numerically the percentage of variance due to heritable factors.

(3) A provisional examination of the existing data shows it to be unlikely that more than 5 per cent. of the variance of the physical measurements of man is due to non-heritable causes. Of other features for which the data is at present insufficient, it would be wisest to judge by comparison of the known facts with those of the physical measurements.

It should not be forgotten that 5 per cent. of the variance would allow considerable scope for the action of environment in individual cases. It amounts to ·338 sq. ins., so that the standard deviation if all else were constant would be ·58 ins. We may put this another way by saying that any factor producing 5 per cent. of the variance is correlated with stature with a coefficient as high as ·224. I am not aware that any environment correlation has ever been established which is more than one-fifth of this amount. So that if there really is a residue of 2 per cent. or 5 per cent. of the variance due to non-heritable causes, there is no evidence of any single environmental cause of sufficient importance to account for the whole of this residue.

In conclusion it is right that I should express my deep sense of gratitude to the Eugenics Education Society, who have most generously assisted me throughout; and in particular to Major Leonard Darwin whose continual kindness and encouragement has enabled me to carry through the work.