PREFATORY NOTE TO CHAPTER II

Any attempt to investigate the relative significance of nature and nurture through modern avenues of investigation is almost certain to lead to more or less elaborate statistical treatment of the numerical data that are assembled. While the science of statistics has within recent years supplied the investigator with various novel weapons of attack, they are often two-edged weapons, whose use is not without hazard to the user. The purpose of this chapter is to describe some of these statistical hazards.

At the request of the Yearbook Committee, Miss Burks has drawn up the presentation of the hazards which constitutes the first part of Chapter II. Her presentation was then referred to Professor Kelley, whose interest in this particular matter is attested by his *Influence of Nurture upon Native Differences,* with the request that he make such supplementary comments and criticisms as he deemed wise. Professor Kelley's contribution constitutes the second part of the chapter.

Readers who have no familiarity with modern statistical methods will probably find portions of this chapter too technical for easy perusal; however, we suggest that they will be repaid if they take the trouble to skim the chapter, omitting the technical portions, because many of the fallacies described therein have been altogether too prevalent in supposedly scientific, not to mention popular, discussions of heredity and environment. We shall never get anywhere in such discussions unless we are perfectly clear as to the fallacies that are likely to beset our thinking.—*Editor*.

CHAPTER II

STATISTICAL HAZARDS IN NATURE-NURTURE INVESTIGATIONS

BARBARA STODDARD BURKS
AND
TRUMAN L. KELLEY
Stanford University, Palo Alto, California

I. CHIEF CONSIDERATIONS

1. Selection

'Selection' is given first place among the hazards because it is so persistent, so widespread, and often so hard to recognize. A practical definition of selection as used here would be: the systematic operation of one or more factors that prevent a group of individuals from being what they are assumed to be. It is found, for example, in attempting to determine how much native difference exists between the mental levels of various races, that in the higher school grades negro children are closer to the level of white children than is the case in the lower grades. On the face of it, this might appear to mean that schooling had wiped out the early difference between negro and white children. If the white and negro children in the higher grades were typical of children of their age, this would indeed be the case. But if it turns out that only the ablest negroes continue at school, it may be that nurture has had no effect at all in narrowing the gap between the abilities of the chosen samples of the two races.

More subtle selective factors may sometimes be at work. If, in an investigation of the mental resemblance between parents and children, 40 percent of the families approached refused to lend themselves as subjects for the experiment, it is evidently possible that the families refusing coöperation might be those in which the children resembled the parents least. Bright parents might be ashamed of children less bright than themselves, and the occasional dull parents who have bright children might shrink from exposing their relative backwardness. Many other examples of selective errors could be cited. The scientific worker must be constantly alert to avoid them.

2. Inextricable Causes

There are many types of study in which this hazard is inherent. Under some conditions the only way to obviate the difficulties is to find a new approach to the problem that will extricate the 'causes' by experimental means.

Let us consider for a moment the correlations of .40 to .50 between the intelligence of siblings or between that of parents and offspring which have been reported by many different investigators. To what are these due? "To nature, to similar germ plasm," answers the hereditarian. "To nurture, to the molding influences of home training and similar educational opportunities," answers the environmentalist.

Either answer is consistent with the observed facts, yet neither answer can be established through the facts immediately at hand. The hypothesis that family resemblance may be due to the *combined* forces of nature and nurture could also explain the observed facts. It therefore behooves us to defer interpretaton until data from studies using different methods of attack untangle the real causes of family resemblance.

Analogous situations could be found in nearly every phase of the field under discussion. Are the correlations of .60 to .80 usually found between intelligence and school achievement due to an influence of intelligence upon achievement, or to an influence of schooling upon intelligence? Burt,[1] arguing from a regression equation for predicting mental age from educational age and other variables, concludes that mental ability as measured on his revision of the Binet Test is the product of schooling to the extent of about 50 percent. Courtis,[1] on the other hand, arguing from a regression equation for predicting educational age from mental age and other variables, concludes that nearly 90 percent of the pupil's school achievement is conditioned by his intelligence. All question aside as to the validity of the statistical methods by which the actual numerical estimates were obtained, it is an interesting, though somewhat disturbing, fact that two such opposing conclusions can be drawn from sets of very similar data simply through the *a priori* assumptions underlying the reasoning in each case.

---

[1] See the summaries of studies by Burt and by Courtis and others in Part II of this Yearbook.

3. Use of Partial and Multiple Correlation

There has been a tendency to assume that the techniques of partial and multiple correlation offer a precise means for evaluating the relative contributions of certain 'causal' factors to a criterion, as for example 'intelligence' or school achievement. For a more extended discussion of the topic than that given here the reader is referred to an earlier paper[2] by the writer. A few excerpts from the paper may suffice to indicate the type of danger that besets the user of partial and multiple correlation techniques.

We may approach the problem inductively by first examining a hypothetical instance of the misapplication of partial correlation so extreme that it would scarcely be approached by any real situation. We wish to investigate, let us say, the effect of age of entering first grade upon subsequent rate of progress through the grades. We may take for our subjects an unselected group of 12-year-old children. After noting for each child (*a*) the age at which he entered school; (*b*) his present grade; and (*c*) his rate of progress as measured by the average length of time it has taken him to complete each grade, we may find a substantial correlation between (*a*) and (*c*). Such a correlation would be reasonably expected if bright children tend both to enter school very young and to receive extra promotions.

Now let us render "present grade" constant by partialling it out. Then the correlation between age of entrance and time required to complete a grade necessarily falls to negative unity, since we are dealing now with children whose early entrance to first grade is exactly balanced by slow progress (*i.e.*, more than normal time spent in each grade), or whose late entrance to first grade is exactly balanced by rapid progress. The condition imposed that all children shall be in the same grade means that age of entrance must completely determine rate of progress. Yet it does not follow that the relationship between age of entrance and progress is absolute, irrespective of the "effect" of present grade upon progress. Indeed, present grade could have produced no *effect* whatever upon the rate of progress preceding it, since it is itself the direct *resultant* of age of entrance and rate of progress. The application of the partial correlation technique has thus been meaningless in so far as we are interested in getting at any true interdependence of age of entrance and rate of progress.

Next let us examine a less extreme situation. In an attempt to determine how much effect differences in environment have upon the development of children's intelligence we might collect data from an unselected group of families giving the intelligence-test scores of a child

[2] *Jour. Educ. Psych.*, November and December, 1926.

and both his parents, and the family cultural status as measured on a specially devised scale. If the results checked with the trends indicated by various investigators, the correlations computed might not be far from the following:[3]

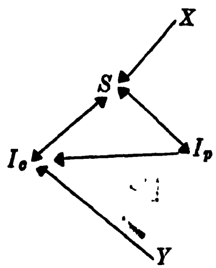
Mental age of mid-parent and intelligence of child......	.60
Mental age of mid-parent and cultural status..........	.77
Intelligence of child and cultural status...............	.48

Using these hypothetical figures, let us calculate the partial correlation between intelligence scores of mid-parent and child. In current terminology we "hold constant" the cultural status, or "eliminate the effect" of cultural status.

Substituting in Yule's formula,

$$r_{12.3} = \frac{r_{12} - r_{13}\, r_{23}}{\sqrt{(1 - r^2_{13})(1 - r^2_{23})}} = .42.$$

Should we then be justified in concluding that the real relationship between the intelligence of parents and their children, after similarities induced by similar cultural surroundings had been discounted, was measured by a correlation coefficient of only .42? Surely not, because we know that by the nature of partial correlation we have eliminated not only those portions of parents' and children's intelligence that may *result* from differences in cultural status, but those portions which *contribute* to cultural status as well. The situation is represented diagrammatically below. I_p represents intelligence of mid-parent; I_c, intelli-



gence of child; S, cultural status; X, factors other than I_p or I_c contributing to S; and Y, combinations of genes not showing in the measurement of I_p contributing to I_c. It is readily seen that in applying the partial correlation formula to this situation we have "partialled out" too much, and that *in any study of causation we are partialling out*

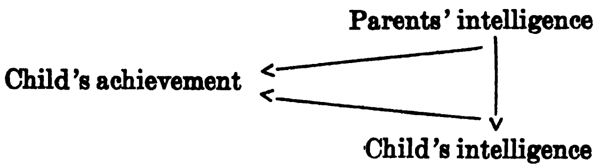
[3] The correlation figures have been changed from the hypothetical values appearing in the quoted article to actual values (corrected for attenuation) computed later from experimental data collected by the writer.

too much when we render constant factors which may in part or in whole be caused by either of the two factors whose true relationship is to be measured, or by still other unmeasured remote causes which also affect either of the two isolated factors.

The same generalization, of course, applies to studies of causation . . . in which mutiple correlations are computed with different factors under investigation successively dropped out to see how much each 'contributes' toward estimating the criterion.

The question . . . . arises as to whether partial or multiple correlation can ever be fully defended as instruments in the study of causation. Surely, as Kelley has pointed out,[4] there is no more justification for inferring causation in partial correlation than in raw correlation. Nevertheless, there are situations in which a variable is indisputably a cause rather than an effect. In one obvious type of situation, chronological age is such a variable. The many studies which have employed partial correlation technique to eliminate the contribution of maturity to correlated measures are apparently on safe ground, provided they have published the precise age ranges of their subjects. However, the caution sounded by Yule[5] in his chapters on partial correlation and normal correlation should not be lost sight of: namely, that except in the relatively rare situation of normal correlation, the partial coefficient cannot be taken as a measure of the correlation between variable 1 and 2 for every constant value of variable 3, but only as a sort of average correlation.

An additional consideration not dealt with in the earlier article suggests a further limitation of the conditions under which the partial correlation is a valid measure of causation. To justify its use as such a measure it is not sufficient merely that the factor partialled out should stand in the relation of cause to the two variables whose interdependence is to be measured. If the variable partialled out is entirely irrelevant to the problem at hand (as we took chronological age to be in our example) the partial correlation does give us what we seek. However, what would be the proper procedure to apply if we had a situation represented by the diagram shown herewith?



[4] *In* Rietz, H. L., et. al., *Handbook of Mathematical Statistics,* 1924, pp. 139 ff.

[5] *Introduction to the Theory of Statistics,* 1924.

The variable to be partialled out, parents' intelligence, is by hypothesis a cause and is not itself caused by any of the other variables entering the problem. The arrows in the diagram indicate the directions which the influences of cause to effect are assumed to take. Since teachers often attribute good class work in part to home training received from superior parents, and likewise attribute poor class work to the lack of encouragement received from unintelligent parents, it would be pertinent to establish, if possible, the relationship between the child's school achievement and his intelligence when the influence of parental intelligence is eliminated.

If we follow the obvious procedure of partialling out parental intelligence, we indeed succeed in eliminating all effect of parental intelligence. But here, as in the first illustration used, we have partialled out more than we should, for the *whole* of the child's intelligence, including that part which can be predicted from parents' intelligence as well as the parts that are due to all other conditioning factors, properly belongs to our problem. We are interested in the contribution made to school achievement by intelligence of a normal range of variability rather than by the narrow band of intelligence that would be represented by children whose parents' intelligence was a constant. The partial-correlation technique has made a clean sweep of parental intelligence. But the influence of parental intelligence that affects achievement indirectly *via* heredity (*i.e.*, *via* the child's intelligence) should stay; only the direct influence should go. Thus, the partial-correlation technique is inadequate to this situation. Obviously, it is inadequate to any other situation of this type.

4. Partial Regression Equations

The use of partial regression equations in studies of causation is subject to some of the same dangers already mentioned under (2) and (3). Causes and effects are likely to be confused or interchanged if there is not some *a priori* basis outside the data themselves for defining which variables are causes and which effects. When interpreting the significance of regression coefficients, even though a rational hypothesis regarding the direction of causal influence to effects has been made, still another type of error sometimes occurs. It is commonly assumed that the regression weights of the so-called 'independent' variables in the regression equations are directly proportional to the actual contributions of those variables to our prophecy of a criterion or 'dependent' variable. It has frequently been said of such a type equation as the following,

$$X = .24A + .61B + .15C,$$

(where *X, A, B,* and *C* are measured in standard scores), that *X* consists of 24 parts *A* to 61 parts *B* to 15 parts *C*; or that in estimating *X, A, B,* and *C* contribute in the proportions 24, 61, and 15. It will now be shown, by means of a simple numerical example, that such an interpretation is not justified, and that the regression coefficients, or 'weights,' can be conceived as conversion factors for putting our independent variables into the same units as those employed in measuring the criterion.

Let the variability in the four variables *X, A, B,* and *C* be due to a number of uncorrelated variable factors as follows:

$$\begin{aligned} X &= a + b + c + d + e + f \\ A &= a \\ B &= b + c \\ C &= d + e + f \end{aligned}$$

We may assume without loss of generality that the variables and their component factors are measured as deviations from their own means, and that the component variables have standard deviations which are all equal to one. We can then easily compute the intercorrelations between *X, A, B,* and *C* by the method which has frequently been used by Spearman and by Thomson. The *r* in each case is $\frac{N}{\sqrt{N' N''}}$ where N is the number of factors common to two variables, and N′ and N″ are the total number of factors in each of the respective variables.[6]

$$\text{Thus } r_{AX} = \frac{1}{\sqrt{6}} = .4082, \text{ etc.}$$

Applying the formula to all the pairs of variables, we get this table of correlations:

	X	A	B	C
X	1	.4082	.5774	.7071
A	.4082	1	0	0
B	.5774	0	1	0
C	.7071	0	0	1

[6] A simple derivation of this procedure was presented by the writer in the article mentioned.

The intercorrelations have been tabled thus because the matrix of the determinant Δ which is used in computing multiple correlations and multiple regression coefficients is composed of the intercorrelations arranged in this way.

Computing $R_{X.ABC}$ by the ordinary formula, we find R equal to unity. This, of course, might have been foreseen, since all the factors in *X* are contained either in *A, B,* or *C.*

Now, computing the regression equation for estimating *X,* we get

$$\frac{\bar{X}}{\sqrt{6}} = .4082\,\frac{A}{1} + .5774\,\frac{B}{\sqrt{2}} + .7071\,\frac{C}{\sqrt{3}}$$

where X, A, B, and C are all measured as deviations from their own means. Thus, we see that in the simple case in which *A, B,* and *C* are independent of one another the weights to be attached to the standard scores of each to predict *X* are equal simply to the correlations of each with *X.*

This does not mean, however, that *X* is composed of *A, B,* and *C* in the proportions .4082, .5774, and .7071; nor even, what might seem more reasonable, that proportions .4082 of *A,* .5774 of *B,* and .7071 of *C* summed together give a composite that is equal to *X.* By hypothesis *all* of *A, B,* and *C* must be summed to give a composite equal to *X.* If we reduce our regression formula above, we see that condition fully met; for, multiplying the left and right members of the equation by $\sqrt{6}$, we get $\bar{X} = A + B + C$, in which the scores of all the variables are measured in the same system of units.

Of course we seldom get results of such simplicity in actual practice because it is only in an artificial situation that we could find a group of variables whose gross scores were all measured in the same units. Even a group of variables which *apparently* employ the same units often suffers from Professor Aikins' "jingle fallacy," as for example in Burt's much buffeted regression equation for predicting Binet mental age. Achievement age, "reasoning" age, chronological age, and Binet mental age were all measured in *years.* But the underlying common trait was *level of intelligence,* and the four variables expressed it in four sets of units which were rendered comparable only by the conversion factors of the regression equation.

A partial analogy from another science may serve to clarify this point. Through the chemical association of sodium and chlorine, ordinary table salt is obtained. What will be the result in terms of salt if we combine a large number of sodium atoms, say 26×10^{18}, with a given volume of chlorine gas, say .48 c.c.?

$$26 \times 10^{18} \text{ sodium atoms} + .48 \text{ c.c. of chlorine} = ?$$

It is obviously impossible to tell, unless the amounts of both sodium and chlorine can be expressed in common units which can be used also to express the salt product. Knowing the weight of an atom of sodium and the specific gravity of chlorine, we can express both in terms of milligrams, thus obtaining the equation for 'predicting' table salt as follows:

$$1.0 \text{ mg. sodium} + 1.51 \text{ mg. chlorine} = 2.51 \text{ mg. salt.}$$

Since regression coefficients are not valid to ascribe relative efficacy to different variables in estimating a criterion, the question arises whether some other device might not be. Since we are dealing with deviations from means (not with absolute levels, *i.e.*, total scores), we are justified in holding that a variable is effective in estimating a criterion in proportion to the amount of variance remaining in the criterion when all other factors are held constant but the variable in question varies as much as before.

In our first example, this is very easy to determine in the case of *A*, *B*, and *C*, since they are independent of one another.

$$\sigma^2_{X.A} = \sigma_X^2 \left(1 - r^2_{XA}\right)$$

The holding of *A* constant thus reduces the full value σ^2_X by a proportion r^2_{XA}.

Proceeding similarly with *B* and *C*, we find for the relative efficacy of the variables in estimating *X*,

$$\begin{aligned} r^2_{AX} &= (.4082)^2 = .1667 \\ r^2_{BX} &= (.5774)^2 = .3333 \\ r^2_{CX} &= (.7071)^2 = .5000 \\ & \qquad\qquad\quad \overline{1.0000} \end{aligned}$$

It thus appears that the *squares* of the regression coefficients (which in this case are also the squares of the correlation coefficients) are proportional to the contributions to estimate. This, like some of our previous results, also might have been predicted, since the fractional expressions for the values of the r's squared are $\frac{1}{6}$, $\frac{1}{3}$, and $\frac{1}{2}$, respectively, and are at once seen to be equal to the proportional number of factors in *X* which are contributed by *A, B,* and *C.*

Sewell Wright[7] has shown that in the general case in which the 'causal' variables are correlated with one another, instead of uncorrelated, the squares of the regression coefficients (which he terms "path coefficients") still represent the proportional contributions made by the different variables to estimating the criterion. The reader is referred to Wright's articles in which the use of these path coefficients is explained. Wright's caution that his method offers no basis for assuming causal relationships, though providing the means for evaluating numerically causal relationships already known to exist, should be borne in mind whenever the technique is used.

5. Nygaard's "Percentage Equivalent for the Coefficient of Correlation"

In a recent article by Nygaard[8] a statistical hazard has been pointed out which concerns the numerical interpretation of the correlation coefficient. But in attempting to develop a method by which a "direct and understandable interpretation may be made of the amount of relationship indicated by a coefficient of correlation" Nygaard has unwittingly fallen into another pitfall.

Assuming (1) that trait *C* depends for its value entirely upon traits *A* and *B,* (2) that *A* and *B* are uncorrelated, and (3) that a weight, k, of *A,* and a weight, h, of *B* combine to give *C,* the author makes the statement that "the ratio of dependence, or percentage of dependence, of *C* upon *A,* will be $\frac{k}{k+h}$, and upon *B,* $\frac{h}{k+h}$."

This statement would be perfectly true if *A* and *B* were measured in identical units, but not otherwise. If John's first helping consisted of one quarter of a pie, and his second helping consisted of one sixth of a pie, his entire dessert would have a "percentage of dependence" upon the first helping of $\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{6}} = \frac{3}{5} = 80$

[7] *Jour. Agric. Research,* 20:1921, 557-585; also *Genetics,* 8:1923, 238-255.
[8] *Jour. Educ. Psychol.,* February, 1926.

percent, and upon the second helping of $\frac{1/6}{1/4 + 1/6} = 2/5 = 20$ percent. But if John's dinner consisted of a stick of celery and half of a fried chicken, the mistake would be immediately obvious if we attributed $\frac{1}{1 + 1/2} = 2/3$ to the celery and $\frac{1/2}{1 + 1/2} = 1/3$ to the chicken.

Since it is almost never that the variables used in predicting a criterion are measured in the same units, the formulas for "percentage equivalents" derived by Nygaard have a very much narrower applicability than he believed them to have.

6. Correlations in Populations of Various Ranges

In view of the numerous discussions that have appeared in the literature regarding the dependency of correlation values upon the range of talent of the populations from which they are derived, it is not necessary to treat this topic at length here. Yet in spite of the comments and formulas for correction that have been published, the fact that in experimental work comparisons between test groups and control groups are occasionally still made without taking any account of possible differences in range indicates that the importance of this hazard is not universally appreciated. Great care is sometimes taken to see that two populations differ by a measurable amount in a 'controlled' factor, such as social status, health, intelligence, etc., yet the control and experimental subjects may be chosen in such a way that a very significant difference obtains in the range of the ability or talent investigated. Sometimes, also, data collected by different experiments may fail to agree simply because the range of the subjects is different in each case. Under such conditions conclusions may be drawn that are conflicting, misleading, or false.

Kelley[9] has tabled the values that correlation coefficients between two variables would take as the range of one variable was extended. If the original r is .60 for example, it becomes .707 if the ratio of the original S.D. to the S.D. for the extended range is .75; the r becomes .832 if the ratio is .50; it becomes .949 if the ratio is .25, and .991 if the ratio is .10, etc.

[9] *Statistical Method*, 1924, p. 225.

7. Incommensurability of Results from Different Tests

There has been a more or less prevalent tendency for investigators to treat test results as though the *names* of the tests accurately defined the functions measured, and as though the scores on any two tests—even if they employ rather different material—are comparable provided only the tests are called by the same name. Thus, for example, we find studies that purport to measure the effect of language handicap on verbal intelligence tests scores by comparing the mental ages of foreign children earned on verbal and on non-verbal intelligence tests. The mental ages of children of certain low-testing nationalities commonly turn out to be closer to the norms of American children when measured on non-verbal tests than when measured on verbal ones. But in as much as verbal and non-verbal test scores, even for American children, seldom correlate with one another higher than .6 or .7, it is obvious that, although both types of tests are called 'intelligence' tests, they each measure about as much not held in common as they measure of what is held in common. Hence, it is not legitimate to infer from such data alone that language handicap accounts for the low scores of the foreign children on verbal tests. It would be only a little less defensible to argue that because certain national groups averaged close to American norms in some such trait as height (which has been shown to have a slight correlation with intelligence), their deficiency on intelligence tests was therefore proved to be due to language handicap.

8. Spurious Index Correlation and Spurious Mutual Correlation with Age

Two sources of error which are often overlooked are discussed by Thomson and Pintner.[10] The first of these, spurious index correlation, results when the paired scores of a correlation table are divided by the same figure, provided this figure varies from pair to pair, and provided the resulting scores show a negative correlation with the figures by which the original paired scores were divided. This situation is quite commonly met when I.Q.'s are computed for mental tests of the kind *for which I.Q. and*

[10] *Jour. Educ. Psych.*, Oct., 1924.

chronological age are negatively correlated. The correlation between I.Q.'s on two such tests is higher than the value representing the real similarity in the functions measured by the tests, owing to the fact that each pair of scores has undergone distinctive treatment.

The other type of error discussed by Thomson and Pintner is that of spurious mutual correlation with age. In the words of the authors: "It is quite possible for two tests to have no organic connection with one another, and yet for the M.A. found by either to correlate highly with chronological age up to even .7 in extreme cases. In such a case there might be no correlation between I.Q.'s, and yet, if the cases were spread well over a long range of chronological age, there might be a very high correlation of M. A.'s between the tests." A striking example of the same type of spurious correlation is that between M.A. and height if a wide age range of children is used. This correlation nearly disappears if children only of a single age group are retained.

It seems appropriate to mention also the type of misinterpretation that is likely to result if I.Q.'s and E.Q.'s are compared on various tests that show *different correlations with chronological age.* It is often assumed that if a child is working 'up to capacity,' his E.Q. will equal his I.Q. Often, likewise, the supposed evenness or unevenness of his various abilities is gauged by the fluctuation in his E.Q.'s for different school subjects. For very rough practical purposes, this procedure may suffice. But it would be perfectly possible for a pupil who had the same percentile ranking for his age in intelligence, reading, and handwriting, for example, to have very different 'quotients' in these traits, simply because the traits showed different amounts of overlapping with chronological age.

The 'standard score' has been proposed and used in preference to the E.Q. by some writers. This gives a pupil's deviation from his age norm in terms of the variability of his age group and would seem to be the most meaningful measure obtainable, providing real age norms are secured. All too often, however, the experimental literature provides instances in which the mean score for a year age-range is used as the norm for a chronological age span of twelve months. Such a method as the one advocated by De Voss[11]

[11] In *Genetic Studies of Genius*, Vol. I, 1925, Chapter XII.

for obtaining an *achievement profile* in terms of 'standard scores' penalizes the children who have just had a birthday, and gives an undue advantage to those closely approaching a birthday. Moreover, these penalties and advantages have different amounts for different tests of an achievement battery, depending upon the correlations of the separate tests with chronological age. The writer has calculated that for certain pairs of tests from the Stanford Achievement battery, discrepancies greater than a standard deviation of the score distribution would occur in children whose *true ability* was the same on the two tests, if these children were measured by norms nearly six months away from their actual ages. The only help for this difficulty would be to secure separate standard score norms for each month of age increment, instead of for year increments only.

9. Confusion Between Variability and Absolute Level

Occasionally, one encounters statements like this: "The ultimate achievement of any given individual is due to his original ability, probably to the extent of 60 to 90 percent, and to actual differences in opportunity or external circumstances only to the extent of 10 to 40 percent.[12] The meaning of such statements is usually far from clear. When we say that an ability owes so much to nature and so much to nurture, do we generally have in mind:

(1) Level of ability of every human being (which the quotation immediately above implies)?
(2) Human level of ability on the average?
(3) Deviations of every human being from the mean of the general population?
(4) Differences among human beings on the average?

Of these four possibilities, (1) and (3) contradict the observed facts, which are that nature and nurture are variable influences by no means perfectly correlated with one another. Consequently, the level of ability or the deviation in ability of different individuals is determined now by one proportion of nature to nurture and now by another proportion.

[12] Starch, D., *Educational Psychology*, 1919, p. 94.

The second possibility would perhaps offer a pertinent line of investigation if any way could be found to carry it out. The writer has never seen any adequate means proposed for making such a determination, however, and strongly doubts whether adequate means will ever be available. To do so would apparently require a knowledge of the effect of nature alone without the aid of nurture, and the effect of nurture alone without the aid of nature. The problem would thus break down, since no development whatever would be possible without the contributions of both nature and nurture.

There remains, then, the fourth possibility as the only practicable and unambiguous concept to use. When the relative contributions of nature and nurture are discussed, no doubt should be left in the mind of readers that contributions to *variability* are being considered.

This comment does not imply, of course, that pertinent and valuable studies cannot be made of the *direct* effects of a variety of influences upon abilities. It is important to know how many points hookworm or an illiterate home can depress the I.Q. or how many points a certain number of hours of study per week or a certain brand of textbook can raise the E.Q. Such investigations probably have more practical bearing than the classical type of study on the proportional contributions of nature and nurture, and indeed constitute the greater bulk of the chapters appearing in this *Yearbook.*

10. Ambiguity of Correlations Between Averages

It is not uncommon to encounter material which presents the correlations obtaining between average scores for a number of groups, each one of which contains a large number of individuals. Bagley, for example, presents correlations of the average Army Alpha ratings of men from the different states matched against average ratings of the states on various educational influences, such as schools, magazine circulation, etc. The correlation coefficients are extremely high—chiefly above .90. These results do not mean that the relationship between intelligence and schooling or magazine reading is as high as .90. On the contrary, studies reporting correlations between the intelligence of *individuals* and the number of years of schooling received by them generally yield coefficients around .60. The correlation coefficients become inflated when averages are employed because a great many factors that ordinarily keep the correlation between intelligence and education from being perfect cancel out (*i.e.*, they are approximately the same, taking the averages of entire states). Consequently, these 'inflated' coefficients cannot be interpreted as ordinary correlation coefficients can be. In fact, it is almost impossible in most cases to give any definite interpretation to them whatever.

11. Manipulation of Scores Without Providing Measures of Group Dispersion

Few things are more exasperating to investigators who wish to make use of previous work in their field than to come upon studies which present increases in certain test scores in terms of points or of percents without indicating the significance of these increases in terms of group variability. Do 10 points or 35 percent average increase on card-sorting scores and 10 points or 35 percent average increase on spelling scores mean the same thing or two different things? In one case such an increase might be equivalent to a jump from the 25th to the 75th percentile of 'unselected' ability, and in the other case, equivalent to a jump only from the 50th to the 51st percentile. It is evident that some measure of group dispersion should always be given in connection with reports of score improvement, and that the nature of the group having such dispersion should be very carefully defined.

12. Probable Errors when Individuals Are Used More than Once in a Scatter

This hazard is met when coefficients of family resemblance are computed in which several offspring are correlated against the same parents or several siblings per family are paired in all possible ways. Owing to the cumulation of errors of sampling when any score enters a correlation table more than once, the probable error of the resulting correlation coefficient is greater than the value given by the ordinary P.E. formula (in which N is taken to be the actual number of entries in the correlation table). If most of the scores enter the table a considerable number of times, the ordinary P.E. formula may be quite inapplicable. Formulas have been derived for determining the probable errors of coefficients so obtained when each score enters the table the same number of times.[18] Unfortunately, the occasion does not often arise when cases available for use can conveniently enter a table the same number of times, if the number of times is greater than one. When it is important to have an accurate determination of the P.E. of r (and it usually is important), either it should be shown empirically that to use the lower or upper limit of N yields approximately the same P.E., or enough cases should be discarded so that every score enters the correlation table the same number of times. If the former can be done, the lower limit of N would, of course, be the number of entries that could be made without using the same case twice. The upper limit would be the total number of entries in the correlation scatter.

13. Assumed Cumulative Effect of Environment

A pageful of citations could be presented in which some statement such as this is made: "It is fair to assume that the longer environment acts, the greater is its effect." By the use of this basic assumption, elaborate 'proofs' are sometimes built up to show that environment can or cannot account for such and such observed facts. The assumption may or may not be true; again, it may be true under some conditions, false under others; it is far from axiomatic. Thus, in some situations it would seem at least as reasonable to postulate that environment quickly accomplishes its maximal effect, and if constant thereafter, is powerless further to add or detract.

14. Over-Simplification of the Mendelian Theory

The attempts made during the last twenty years to fit to the observed facts of family resemblance in mental traits a theory which would account for the mechanics of hereditary transmission are probably premonitions of immensely important and serviceable developments in the eugenics of the future. It may truthfully be said, however, that these attempts as yet have done little more than scratch the surface of the problem, and that the chief value of

[18] Smith, K., *Biometrika*, 14:1922, 1-22.

some of them is in stimulating controversies that may result in further work along more productive lines.

The studies that have aroused so much criticism are of the type which have investigated family resemblance on some highly complex, continuously graded trait like intelligence, insanity,[14] nomadism, or musical talent, have attempted to categorize it as present or absent, and have made counts of its appearance in several generations to see whether this can be made to fit any of the simple Mendelian ratios.

The surprising thing about the results of such experiments is that the traits often *appear to behave almost as though they actually were Mendelian unit characters,* although 'unit character' and 'continuously graded trait' would seem to constitute an emphatic contradiction in terms. The agreement between experimental data and simple Mendelian ratios is never perfect, but is sometimes well within errors of sampling, and is often marked enough to make a simple Mendelian hypothesis appear quite reasonable were it not for the contradiction just noted.

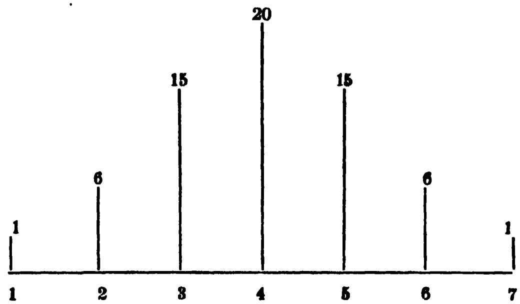
As a matter of interest, I have performed some calculations to ascertain whether a hypothesis of cumulative Mendelian genes (which certainly would be more in accordance with observed distributions of mental traits than a 'unit character' hypothesis) might not be made to give as good approximations to simple Mendelian ratios as those reported in investigations of mental traits by Goddard, Davenport, Hurst, Rosanoff and Orr, Cobb, Peters, *et al.* It is not asserted that the hypothesis utilized in the following calculations is a correct one; it is presented merely to show that a more probable mode of hereditary transmission than that considered by some of the 'simple Mendelian' advocates can account for the ratios of two arbitrarily designated types of offspring about as well as can the over-simplified theory, and at the same time account for trait distributions in a general population far better.

Assume (1) that intelligence is due to three cumulative pairs of genes, Aa, Bb, Cc, showing neither dominance nor epistacy, (2) that the two phases of these genes have equal incidence in the population, (3) that all genes make equal contributions to the trait

[14] Strangely enough, all forms of insanity have sometimes been lumped together.

intelligence, and (4) that mating is at random. (The latter assumption was probably unnecessary to make as far as final 'Mendelian ratios' were concerned; it was made simply because the calculations would otherwise have become very unwieldy.)

The resulting population showed a distribution of seven grades of intelligence in the proportions shown in the following figure.



Assume now that the two lowest grades of intelligence correspond to 'feeble-mindedness' and the five other grades to 'normality.' The numbers of offspring of various genetic formulas were computed for parent combinations covering the entire gamut of random mating, assuming four offspring to the mating; and the probability was allowed for that in a definite proportion of families for which the expectation would be three 'normals' to one 'feeble-minded,' half-and-half, etc., all four offspring would turn out to be normal or feeble-minded. Goddard's scheme, described in his book *Feeble-Mindedness,* was used for determining whether or not parent 'formulas' in certain matings should be called simplex, or indeterminate for 'normality.'[15]

Goddard's notation for designating types of mating is used, and a few figures from his study on *Feeble-Mindedness* are compared with our hypothetical ones.

NN designates the duplex normal individual, NF the simplex normal, and FF the nulliplex, or feeble-minded.

---

[15] A normal parent is called 'simplex' if his mating results in any feeble-minded offspring, and 'indeterminate' if his mating results in no feeble-minded offspring, since in the latter case he might conceivably be either simplex or duplex.

Type of Mating	Offspring Feeble-Minded (Percent)	Offspring Normal (Percent)
FF — FF		
Artificial population	81.6	18.4
Goddard	98.7	1.3
Simple Mendelian expectation	100.0	0.0
FF — NF		
Artificial population	47.3	52.7
Goddard	57.3	42.7
Simple Mendelian expectation	50.0	50.0
NF — NF		
Artificial population	28.1	71.9
Goddard	31.9	68.1
Simple Mendelian expectation	25.0	75.0

The only place where the ratios of 'feeble-minded' and 'normal' offspring in the writer's artificial population differ radically from Goddard's factual data is in the FF — FF type of mating where the artificial parents seem to produce too many normal offspring. If the number of genes and line of demarcation between normal and feeble-minded postulated for the artificial population should be altered by trial, it is probable that better agreement could be found at this point. The thing to be emphasized here is that the right combination of two feeble-minded parents can occasionally produce normal offspring if cumulative genes account for intelligence, whereas this would not be possible by a simple Mendelian theory. Goddard's 1.3 percent normal offspring, which he attempted to explain through occasional inaccuracies in the estimated intelligence of individuals entering his count, may thus be the necessary consequence of the actual conditions of inheritance.

II. Supplementary Considerations

It might not be out of place to suggest what a few needed avenues of approach to the nature-nurture problem appear to be, and to indicate as succinctly as possible some facts and possibilities that could well be kept in mind in planning new work.

1. Further Studies on Specific Effects of Specific Influences

We need more studies of the type which this Yearbook contains: accurate determinations of the effect of diseases or of various physical conditions upon mental development and the ability to do school work; careful evaluation of the effect of specific elements of home or school environment as differentiated from general excellence of either; measurement of racial differences with the possible effects of environment eliminated through experimental control; studies, employing rigorously matched control groups, to measure the week-by-week improvement on mental tests and achievement tests of foreign-speaking subjects as they learn English; the effect of subtle factors of personality upon the I.Q. and upon the use to which the I.Q. is put, etc. Such investigations would offer a life time of work to a corps of the best research workers that psychology, biology, education, and sociology could produce.

Of all the influences enumerated above, perhaps that of home environment is the most important to understand and evaluate. Professor Terman has proposed an experiment which would give very clear and conclusive data upon the effect of home environment on the I.Q. and would be entirely feasible, though very costly. Briefly, the experiment is: "From several hundred families of the grade whose offspring ordinarily yield a mean I.Q. of 80 (say, families of low-grade unskilled laborers), take 500 children as soon as they are born, and, after subjecting them to ten years of superior educational and cultural influence, compare their I.Q.'s with those of 500 other children from the same families who have not had these advantages. The mean I.Q. difference found would measure the combined effect of the environmental opportunities enjoyed by one group and denied to the other."

Eventually, too, an elaborate synthesis will have to be made. It must be determined whether or not the effects of various factors, singly studied, combine additively, or whether some effects are swallowed up in the presence of greater ones. For example, if poliomyelitis should lower the I.Q. an average of 15 points, and if lethargic encephalitis should lower it 10 points, would subjects who had suffered from both diseases be retarded 25 points, or would all the possible damage have been done by the first disease?

2. Nature-Nurture Contributions to Other Traits

It is probably not putting it much too strongly to say that we know next to nothing about the way character traits, interests, attitudes, ambitions, and special ability traits of all kinds are conditioned. Farnsworth, for example, in a summary appearing in Part II of this *Yearbook,* shows what an unsatisfactory state our knowledge of the inheritance of musical talent is in, and this in spite of the fact that more studies have undoubtedly been made which seek to uncover the determiners of this talent than of any other kind of quality just enumerated.

The *Yearbook* offers in Part II exploratory studies by May and Hartshorne on honesty, and by Anderson on mechanical ability. It is to be hoped that these two chapters may help to stimulate an extensive program of research in the field of personality and of special ability traits. There are many psychologists and practical educators who would put temperament and special bent even beyond I.Q. or E.Q. in importance to the individual and to society.

3. Mechanics of Mental Heredity

If one of our ultimate goals is the accurate control of mental endowment through eugenics, it is not enough to know the proportional contributions of nature and nurture to ability, nor even to know the specific effects of given amounts of various influences, such as parental intelligence, home environment, schooling, etc., upon ability. To be able to predict with any assurance what type or types of offspring to expect from any combination of parents, it is necessary to get in some way at the genetic constitution of the parent generation, to find out whether or not mental traits are transmitted by Mendelian factors, and if so, through interminable research, slowly to identify these factors, locate them in the chromosomes, and determine their degrees of dominance.

Practically the only studies which have approached mental heredity from the point of view of genetics are of the type mentioned in an earlier section of this paper where it was pointed out that a view of Mendelian heredity far too narrow to fit the observed facts had been employed.

However, a few theoretical studies deserve mention in which the attempt has been made to develop broad generalized hypotheses of Mendelian inheritance in 'continuous' traits and to compare the 'expected' correlations between relatives with the correlations found experimentally as a check upon the hypotheses. Although these studies have employed data from physical, rather than from mental traits, they supply the cues for investigations of mental heredity.

Pearson,[16] turning in 1904 to the newly rediscovered Mendelian theory, and assuming that a trait was due to N cumulative genes showing perfect dominance, calculated that under random mating the correlation between parent and offspring would be .33. As this value was out of harmony with the coefficients that had been found for parent-child measurements on a large number of continuous traits, Pearson definitely turned away from the theory which he had found inconsistent with experimental results, and never again made any serious attempt to interpret his own data by means of it.

It was subsequently pointed out by Yule,[17] and later by others, that parental correlations higher than .33 and quite consistent with actual values would follow from a generalized Mendelian theory if complete dominance in genes were not assumed.

It remained for Fisher[18] to work out a scheme by which the correlations between parents and offspring, between siblings, and between fathers and mothers are used to infer a coefficient of environment, "*the ratio of the variance* [of an unselected population on a continuous trait] *with environment absolutely uniform to that when difference of environment also makes its contribution.*" Fisher shows that when a trait is due to Mendelian factors, "the effect of dominance is to reduce the fraternal correlation (*i.e.*, its genetic value of one half) to only half the extent to which the parental correlation is reduced," and that this effect "is independent of the relative importance of different factors or of their different degrees of dominance." Making the important assumption that *environment works in a random manner,* thus reducing rather than raising correlations between relatives, and reducing fraternal correlations to the same extent as parent-child correlations, Fisher then utilizes the differences actually found between fraternal and parent-child correlations to distinguish between the effects of dominance and those of environment, finally arriving at the coefficient of environment defined above.

---

[16] *Phil. Trans.,* 203 A, 1904, 53-87.
[17] 1906 Conference on Genetics, *Horticultural Society's Report.*
[18] *Trans. Royal Soc. Edinburgh,* 52:1918, 399-433.

The difficulty of applying Fisher's scheme to family correlations in mental traits is that the effects of environment cannot here be assumed to be random. Demonstrably, environment works here to increase family resemblances rather than to decrease them, and it is probable that this increase is higher for siblings than for parents and offspring. Consequently, Fisher's method could not be applied to problems in mental heredity without fundamental modifications.

Another possibility not considered by Fisher (and as far as the writer knows, not by any other pioneers in the theory of hereditary transmission) is that *environment may have different degrees of influence when the endowment for a given trait is of larger or smaller amount.* It would seem reasonable to suppose that, when a large amount of a trait was present in the genotype, the possibilities of somatic fluctuation would be greater than when only a small amount was present.

This supposition is rendered quite probable in the light of results from a study of parent-offspring correlation by Davenport.[19] Although Davenport's study is concerned with measurements of height, rather than with mental measurements, it provides one of the best series available of family data on a complex trait and is considered here in the absence of adequate data upon any mental trait.

Davenport finds regression almost nil in the offspring of two very tall parents, but quite pronounced in the offspring of two very short parents, and intermediate for 'mid-parent' heights in between. He interprets these results as meaning that tallness is due to the *lack* of inhibiting factors, while shortness is due to inhibiting factors, and that the inhibiting factors are dominant, thus allowing for greater variability in the genotypes of very short parents than of very tall ones. This hypothesis would account for the greater amount of regression of offspring of short parents, but would be inconsistent with the nearly symetrical distribution of height universally found in unselected populations.

If, instead of Davenport's assumptions, a differential effect of environment were postulated, so that the somatic variability of the genotype increased in proportion to the number of inhibiting fac-

[19] *Eugenics Record Office Bull.* No. 18, 1917.

tors present, the phenomena of more variability and greater regression in offspring from short parents could be explained without the untenable condition that the dominant phases of all genes for stature are in the direction of shortness. The greater variability, under the writer's hypothesis, would be due to the fact that individuals from a number of variable genotypes would wander into any given phenotype more frequently and from more distant classes at the short end than at the tall end of the distribution. The offspring would then tend to revert to the original genotypes. Regression of offspring on mid-parent would occur, not because there was any 'urge' within the germplasm to return to mediocrity, but because more genotypes would wander into any phenotype from the direction of the mean than from the direction of the extreme of the population, simply because there *are* more and more individuals, the closer we approach the mean of any normally distributed population.[20] Using the 'artificial population' described earlier in this chapter, the writer computed how much regression in offspring would occur when certain fixed amounts of environmental 'susceptibility' per contributing gene were postulated, and found curves of regression plotted against mid-parent of the same general type as the one reported by Davenport for height—*i.e.*, pronounced regression at one extreme of the distribution, and no regression at the other extreme. It will be interesting, as a future problem, to find out whether data from mental traits also show such characteristic properties of mid-parental regression.

While the general agreement between Davenport's experimental results and the writer's theoretical calculations cannot be said to establish the theory, at least it suggests the expediency of further work in the same direction.

COMMENTS UPON "STATISTICAL HAZARDS IN NATURE-NURTURE INVESTIGATIONS"

By Truman L. Kelley

Miss Burks has rendered an important service in having very clearly pointed out a large number of treacherous pitfalls that lie in wait for the unwary pilgrim treading the narrow and obscure

[20] The writer is indebted to Professor L. L. Burlingame, Stanford University, for this latter suggestion.

path that winds its way between the eternal oaks of nature and the ephemeral shrubs of nurture. It is difficult to meet the request to comment briefly upon the many very pertinent issues raised in the foregoing chapter because most of the hazards of experimental and statistical research are matters of detail. Errors are very insidious in their entry into a study and in their effect upon conclusions drawn. For this reason comment concerned with the larger issues will generally fail to note the critical matters wherein danger lies, and such comments may be but an elaboration of the obvious.

An experimental study with its attendant statistical treatment may be thought of as constituting a logical argument wherein one false step vitiates the subsequent development. It has at times been considered analogous to a chain, the strength of which is only that of the weakest link. This analogy does not always hold, for in some experimental studies there are so many independent and semi-independent lines of evidence that there is not a single connection of premise with conclusion. The chain analogy should be replaced by that of a thread composed of many strands. When a study is of this type, each strand should be subjected to a thorough scrutiny and tested as rigorously as though upon it alone fell the entire weight of the argument. In the process of testing a line of argument one is almost certain to find assumptions that are only approximately true, such as, for example, "The correlation between height and weight is linear," or "the longer nurture acts, the greater is its influence," or "the distribution of 'leadership' in the case of random twelve-year-olds is normal," etc. Every such assumption is undoubtedly a source of error. Not uncommonly we may be sure that the error is not serious, but even so it is far better to continue to note that it has introduced some uncertainty than to dismiss it from mind. If there is a succession of such assumptions in a study, we can clearly see that their total effect will be such as is no longer pictured by the analogy to a chain or to a thread of many strands. An imperfect link in a chain, if stronger than the weakest link, is just as serviceable as though it were without fault, but a succession of weak steps in a scientific argument may terminate in a conclusion far weaker than any of these steps singly, for there may have been correlation between flaws. A chain is held at both ends, but an original psychological experiment should proceed from given data whithersoever the facts truly judged compel. Let us, therefore, adopt another picture. A tower built of cardboard cubes standing one upon another is weaker and less stable than any one of the blocks singly. If the heavy edge of each block is placed to the north, the tower will be much less stable than if the heavy edges are now north, now south, now east, now west. If, in a social science experiment, chance errors now work for and now against a certain conclusion, a certainty is attained which is clearly in excess of that built upon steps, the successive errors of which are correlated. The only safeguard is, first, to attempt to determine carefully the size or nature—chance or biased—of the error in each step, and second, to give careful consideration to the nature and net outcome of the summation of all errors.

Miss Burks' presentation, though admirably calling attention to many sources of error, seems to me not to lay sufficient stress upon chance errors, such as are commonly present in social science investigations. In the section dealing with partial correlation attention is called to the fact that the partial correlation coefficient $r_{12.3}$ is the result of partialing out too much when "we render constant, factors which may be in part or in whole caused by either of the two factors whose true relationship is to be measured, or by still other unmeasured remote causes which also affect either of the two isolated factors." This is true, but so far as presenting a picture between true variables, *i.e.*, between such as have no error factors in them, $r_{12.3}$ is also the result of partialing out too little, for we really should have partialed out the chance factors in the several variables.[1] Thus, if the variables are x_1, x_2, x_3, and if x_1 and x_2 contain, respectively, the chance factors x_4 and x_5, the partial correlation that people generally think they are getting when they get $r_{12.3}$ is $r_{12.345}$. With such measures as we must now frequently deal with, x_4 is ordinarily large with respect to x_1, x_5 large with respect to x_2, so that the numerical value of $r_{12.3}$ may be very different from $r_{12.345}$. I do not object to the point made in Section 3, with reference to partial correlation coefficients, when, as is seldom the case, it is known which variable or variables must

[1] Attention is called to the fact that Miss Burks, in the example she offers, has partialed out the chance factors, inasmuch as the correlations used by her were coefficients corrected for attenuation.—L. M. T.

be the cause and which the effect, but suggest that an even larger source of error in interpretation, that due to chance factors, has been overlooked.

There are two serious difficulties in interpretation of correlation coefficients, and it should be noted that partial correlation coefficients hold no different place in this respect from that held by ordinary total correlation coefficients. One speaks of the score of an individual on the XYZ Intelligence Test as the individual's general intelligence, whereas this is ordinarily in substantial error for two reasons. First, the score has a large chance factor in it, and second, the XYZ Test, in so far as it is not chance, may be a measure of something quite other than intelligence. In Section 3 there is given an illustration wherein one of the variables is "cultural status." I do not know what was the specific measure of cultural status employed, but knowing the difficulty of securing any measure of this type, I venture to suggest that probably about half of this measure is chance,[1] and probably about one-half of the half not chance is something other than cultural status as conceived by the modal American. One cannot partial out cultural status by partialing out a measure, one-fourth of which only, roughly, is entitled to the name. This probably sounds very elementary, as in fact it is, but why it should appear simple when dealing with partial correlation and not equally obvious when dealing with total correlation is a puzzle which must be left to the reader. Is it not clear that one cannot correlate social status with something else by correlating a measure, only one-fourth of which is social status? If the person labelling the social-status measure objects to the statement just made, and states that he explicitly wishes it understood that 'social status,' as he uses the term, means score on his social status scale, the situation is but slightly improved, for (a) it is extreme to incorporate a purely chance factor of large amount into a thing named in such a manner as not to imply chance, and (b) it is a severe, even if not impossible, requirement of a reader to ask him to discard an already established concept and build up another which is perhaps dependent upon the summation of scores or even of hundreds of items. "Social status" has been used merely as an illustration, for every mental measure is open to the same criticism.

[1] See footnote, p. 35.

In brief, though I agree that such interpretations of partial correlation coefficients and regression coefficients (and I wish total correlation coefficients had been included as well) as are criticised in Sections 3 and 4 are unsound, I doubt if such precautions as are suggested are sufficient. Rather—it seems to me—one should not attempt to reach such abstract conclusions. Why cannot experimentalists use measures in many connections; note the properties of each measure as shown by its reliability, as shown by age, race, sex, vocation, and other differences; calculate total, partial, and multiple correlations between these measures, now rich in meaning; and draw conclusions in terms of these specific measures and not in terms of unmeasured abstractions. Surely, we would be in an interminable morass if the relationships found between some hastily conceived "persistence of motives" or "honesty" test had to be accepted as giving us the true relationships of these traits. I believe that the agreement between the modal social concept corresponding to variously named character or mental traits and the traits measured by tests similarly designated will only run from 10 to 70 percent of the variance of the test employed. If I am correct in this, the point here raised is clearly of major importance.

In the discussion of the relative efficacy of different variables in estimating a criterion (Sections 4 and 5) is the statement: "We are justified in holding that a variable is effective in estimating a criterion in proportion to the amount of variance remaining in the criterion when all other variables are held constant, but the variable in question varies as much as before." The variance is equal to the standard deviation squared, so that the statement is equivalent to saying that the argument should be based upon the square of the standard deviation instead of upon its first power or any other power. Personally, I prefer this method to any other, but I would hesitate to say that this procedure is "justified" in any conclusive sense.

Section 13 deals with the cumulative effect of environment. Therein is found a criticism of the hypothesis: "It is fair to assume that the longer environment acts, the greater is its effect." This either is, or might well be, a quotation from the present writer. I have used the principle that the longer environment acts, the greater its effect, but only in connection with growing traits. In fact, I have not taken the environmental influence as proportional to the time through which it has acted, but proportional to the change, as judged by the growth of average individuals, that has taken place in the function during the time through which the environment has acted (see Kelley, *The Influence of Nurture Upon Native Differences,* page 16). This seems to me the most reasonable hypothesis to make with reference to growing traits. There are, of course, other situations wherein environment quickly accomplishes its maximal effect. For example, Jones is introduced to Brown and continues to know him for ten years. The sixty seconds in which the introduction took place may well be more potent than the entire remainder of the ten years in causing Jones to remember the name Brown. Here, however, after the first learning of the name, we no longer have a growing function.

Every hazard noted in the foregoing chapter is real, and unfortunately each has been overlooked time and again. A careful study of this chapter should make explicit, and should therefore lead to an avoidance of, many serious errors that now contaminate studies of heredity and environment.