BIOMETRICS 34, 1–23
March, 1978

A BIOMETRICS INVITED PAPER

Logical, Epistemological and Statistical Aspects of Nature-Nurture Data Interpretation

OSCAR KEMPTHORNE

Statistical Laboratory, Iowa State University, Ames, Iowa 50011, U.S.A.

Summary

In this paper the nature of the reasoning processes applied to the nature-nurture question is discussed in general and with particular reference to mental and behavioral traits. The nature of data analysis and analysis of variance is discussed. Necessarily, the nature of causation is considered. The notion that mere data analysis can establish "real" causation is attacked. Logic of quantitative genetic theory is reviewed briefly. The idea that heritability is meaningful in the human mental and behavioral arena is attacked. The conclusion is that the heredity-IQ controversy has been a "tale full of sound and fury, signifying nothing". To suppose that one can establish effects of an intervention process when it does not *occur in the data is plainly ludicrous. Mere observational studies can easily lead to stupidities, and it is suggested that this has happened in the heredity-IQ arena. The idea that there are racial-genetic differences in mental abilities and behavioral traits of humans is, at best, no more than idle speculation.*

List of Contents

Summary.

1. *Introduction.*
2. *Some Background.*
3. *Intelligence and IQ.*
4. *General Data-Interaction Ideas.*
5. *What is a Viable Notion of Causation?*
6. *Data Analysis and Analysis of Variance.*
7. *Correlation, Causation and Path Analysis.*
8. *Independence.*
9. *Consequences of Falsity of Independence.*
10. *The Case of Random Variables.*

Key words: Nature-nurture; Intelligence; IQ; Data analysis; Causation; Analysis of variance Path analysis; Heritability; Prediction; Observational studies; Intervention studies; Race differences.

11. Is Partition of Variance Sometimes Useful?

12. The "Logic" of Quantitative Genetic Theory.

13. Heritability.

14. The Burt Case.

15. Data Analysis in the Area.

16. Can We Boost IQ?

17. Observational Science in General.

18. Prediction Versus Causation.

19. Concluding Remarks.

1. Introduction

There has been in the past ten years or so a huge amount of discussion and controversy about the role of genes, "nature", and the role of environment, "nurture", in the determination of what has been called intelligence, which in the general setting is "what intelligence tests measure". The thesis of the present paper is that the arena has been littered with poor science in nearly all directions and, in particular, at one pole in the argumentation of the hereditarians who advance the hypothesis of nearly complete genetic determination, and at the opposite pole in the argumentation of the non-hereditarians or the "environmentalists" who advance the hypothesis that genetic causation is negligible. It would take an inordinate amount of space to document this appraisal of the arena, and I shall not attempt to do so, except to state that the names of A. Jensen, Schockley and H. J. Eysenck are now almost household names because of the views they have put forward. Obviously, there are others.

The purpose of the present paper is to present and discuss logical, epistemological and statistical aspects of the whole problem-model-data-inference situation. It is of high relevance that these aspects can be discussed in the context of a wide variety of human or even non-human organismal attributes, and a large portion of the discussion I give can be applied to, say, the functioning of the heart, to physical skills, or to many behavioral attributes of any animal species. I hope and wish the methodology that has been infused in the nature-nurture controversy will be examined *on its own* without bringing in the naturally highly emotional idea of intelligence of humans.

It will be necessary, however, to take up aspects which are intrinsic to discussion of intelligence and the measurement of this supposed attribute. It seems most natural that the whole topic should be discussed in *Biometrics,* because the problem and the modes of attack are almost exemplar of the thrust of biometrical approach.

2. Some Background

The idea that some underlying or intrinsic genetic force was involved in the determination of all human attributes, physical and mental, was held, it would seem, long back in antiquity. The idea that inbreeding was bad for a society goes back similarly; one need only mention the Hammurabic Code of 1792–1750 B.C., items 154, 157 and 158, and Chapter 18 of Leviticus. Attempts at rational qualification and quantification of the role of heredity were initiated, it seems, by the eminent Victorian, Francis Galton (1869) in his book *Hereditary Genius.* This work is surely a remarkable effort of its time; but the intellectual arrogance and the

prejudices of the "enlightened" Victorian gentleman shine through for us who are "really enlightened" to see. I shall not attempt to document this assessment, and merely give two quotations:

> *"The theory of hereditary genius, though usually scouted, has been advocated by a few writers in past as well as in modern times. But I may claim to be the first to treat the subject in a statistical manner, to arrive at numerical results, and to introduce the "law of deviation from an average" into discussions on heredity".*

> *"The method employed is based on the law commonly known to mathematicians as the 'frequency of error'"* (which Galton claims to rest on a perfectly just basis).

It is also interesting to note the following. Galton asks "Is reputation a fair test of natural ability?", and responds "It is the only one I can employ." This arose later with regard to the IQ test, and correlation of IQ and occupational status, which validates (?) IQ.

Galton is surely to be regarded as one of the founders of biometrics which is regarded, justifiably, as a "quantum" leap in human rational effort, with roots going back, I suppose, to Aristotle and the view that the first step in attempting to understand Nature is description and classification. But Galton claims far too much from his data and methods.

The next step in the nature-nurture problem, undoubtedly, was the discovery of Mendelism. At the beginning this was thought to be a fine theory for dichotomous or "multichotomous" qualitative traits, and it was denied that Mendelism should be and could be incorporated in the explanation of so-called continuous variation. I shall not review this in any depth and shall merely mention the names G. Udny Yule and Karl Pearson. The first big step in the breakthrough was surely by W. Weinberg (1908), in work which the scientific world appears to have been unaware of for two or more decades. The main breakthrough was made by R. A. Fisher (1918) in a paper that is remarkably difficult to understand, so much so that it is still under debate. [See for instance Vetta (1976)].

This paper is intrinsic to the nature and history of the IQ controversy. It is, indeed, *the* basis for a considerable portion of the work of the past two decades; this is *often unmentioned,* and perhaps not understood even when mentioned—I do not understand completely myself after perhaps a year of intensive study over the past three decades. The paper attempts and, I believe, successfully, to show that "continuous variation" and "discrete" Mendelism are reconcilable. The topic is not easy and my own book (Kempthorne 1957) may be regarded as having as its major thrust this reconciliation and the use of the reconciliation for the improvement of organic stocks.

Fisher, of course, realized the potential impact of Mendelism and his own ideas on human affairs. He became a leader in the eugenics "movement" and perhaps one third of the first volume of his collected papers deal with eugenics. I have commented on this elsewhere (Kempthorne 1974) and shall merely remark here that Fisher appears to me to be, and naturally so, an individual with a full complement of what I term the prejudices of the Victorian enlightenment. This does not, of course, detract at all from the magnitude and importance of his intellectual effort.

It is correct, I believe, to state that Cyril Burt, who had made *great* contributions to the arena of mental testing and statistical treatment, and data analysis of test measurements, discovered the Fisher work in the early 50's. This led to a paper with a lady, Margaret Howard, whose existence and nature is currently a mystery (Burt and Howard 1956). Burt was a leader of a school and some of his advanced students, in particular A. Jensen and H. J. Eysenck, who are principal protagonists in the current controversy, accepted and propagated the Burt theory, the essential substantive conclusion of which is that the heritability of IQ is of the order of 80%.

The culmination of this path is the argumentation carried on, unfortunately most frequently in the popular press, with numerous substantive and polemical writings. I have to express my view, which space and time considerations prevent me from substantiating, that to a considerable extent the review processes which supposedly scientific journals appear to have followed are strongly defective, and the journals involved are amongst the most prestigious of the world. Furthermore, in my opinion, the downright failure of rational scientific development has not been confined to one pole of the heredity-environment controversy. There is so much garbage (sic) interwoven with objective reasoning, that the reader can only be thoroughly perplexed. There have been "gutter" polemics, emotional and irrational discourse to the extent that the rational portion of the scientific world tends to "switch-off". It has often been said, correctly, that the proper study of mankind is man. The development of knowledge of the nature of human mental abilities is surely one of the most important and ethically valuable pursuits of science. But it cannot or should not be used to justify shoddy scientific work.

The detonator of the controversy was the paper by A. Jensen (1969) in the *Harvard Educational Review,* which was discussed in a succeeding issue, often competently, but unfortunately incompetently also. I shall not attempt to document in detail these *opinions.* The present paper will give a partial basis to them.

3. Intelligence and IQ

While I assert with confidence that the overall process of examining the nature-nurture question is common to a very wide array of scientific questions and problems, and stands or falls *in general,* I deem it necessary to give a few remarks about the particular substantive issues of intelligence and IQ.

It is obvious that we can measure a wide variety of abilities in man, animals and plants. We can measure physical abilities, e.g., running, lifting and withstanding particular sorts of physical stress. Or we can measure a variety of mental abilities, e.g., in humans computing sums of arrays of numbers, solving linear equations and differential equations, recalling a list of words, arranging blocks and so on. We can then, surely, construct a battery of tests. We can apply these tests to a group of individuals, e.g., of age 8, and we can construct from the battery an index of ability, which is standardized to give results which are like a random sample from a normal distribution with mean 100 and standard deviation 15. What name we give to this standardized score is essentially arbitrary. We should use, perhaps, a neutral or "non-loaded" name such as Stanford-Binet score. But this has not been done; instead, the score has been called Intelligence Quotient or IQ. Is this appropriate? We could equally well ask why a particular type of animal should be given the name "dog", or why a particular measurement result should be called "temperature". But it behooves a scientific discipline to have a decent respect for the customs of society in which a word has already acquired a meaning, even if somewhat vague.

It is clear, I suggest, that the mental testing discipline overstepped reasonable bounds in labelling results of the tests it developed as intelligence scores. It is clear, I further suggest, that there is a huge array of mental abilities which humanity needs and values and puts under the general rubric of "intelligence" which do not appear in the batteries that the "intelligence-testers" use. Any ordinary citizen could enumerate many, except for one feature about which any rational human society should be concerned. What is called intelligence testing is now an industry, motivated (but not dominated, I believe) by a self-serving (and natural) profit motive, with all the necessary trade secret apparatuses. The industry is a closed one, which insists that it be its own monitor. It is, of course, performing well and

doing what society needs. Why? Because it examines itself and what it sees is good! So we have the occurrence of SAT testing and of ACT testing, which then determine our "geniuses" and our "idiots". But when we examine these tests even with the limited ability we have of so doing, we infer that there are extremely important aspects of mental functioning *which are given no attention at all.*

Because of the inbuilt secrecy, I cannot do more than make general suggestions. These are that artistic, musical, linguistical, human relational, mechanical (mind-physical coordinative) abilities and others, which I assert to be primarily or even critically mental, are not covered by the standardized tests that are *sold* to our society. So, indeed, "intelligence" is "what intelligence tests measure". But to say this is no more or less than a "cop out" in our modern vernacular.

The reification of the IQ score that has taken place is unfortunate, and the intelligence-testing profession and the intelligence-testing industry must surely take much of the blame. One may well ask of these groups, and of the academic profession, and of business whether they would be prepared to take the IQ test and then be judged as individuals by their scores and even recompensed by an index in which IQ score entered significantly. None of these would be prepared to do so and, I believe, correctly so. One has the story of Poincaré, one of the intellectual giants of all time, who obtained a score in the moron range.

But there is another side of the coin which Jensen and others state. The measurement of various aspects of mental ability is critical for a human society, even for a primitive one which must choose leaders. The *abilities* that IQ tests measure are of considerable importance to *some* of the activities of our society, and it is quite definite that ability to perform well or even adequately in many aspects of our society is associated to some extent with IQ scores. So the individual with a very low IQ score will likely be unable to obtain and hold down a useful and satisfying job. Hence what "intelligence tests measure" may be justifiably regarded as *partially* relevant to the activities and hopes for a satisfying life.

I take the view then that IQ measurements are as valid as any other sort of measurement. The aspects of the overall problem that I shall discuss are aspects common to a wide variety of measurement processes. It is absolutely essential to state this, because there are many individuals who have demonstrated high ability in physical or biological science and who take the view, as I interpret their writings, that it is simply ludicrous to regard measurement and study of mental abilities as falling within the domain of science, likening the arena to astrology and such. I shall not document this assessment. The extreme point of this position is that IQ measurement is mumbo-jumbo. It is a possible and reasonable view that IQ measurement does not measure what the critic regards as intelligence, a notion which he uses but, it would seem, refuses to believe to be susceptible to measurement.

4. General Data-Interaction Ideas

Let us take height on the twenty-first birthday as a trait which interests us. We see in our existent human populations the occurrence of considerable variability. We shall, usually, look at the distribution of height, and may note that, either as obtained or as transformed in some way, we have a distribution which is like the Gaussian distribution, and we shall then have a succinct description that height $\sim N(\mu, \sigma^2)$. We shall be happy with this description, as were early biometricians, for a while, perhaps too long. But we shall eventually be dissatisfied, and we shall ask such questions as (i) Why is the distribution Gaussian? (ii) Why is the mean equal to 69? (iii) Why is the variance equal to 4 (say)? Or we may find that our distribution is like a mixture of two Gaussian distributions. Indeed, there is no possible limitation to what we might find.

So given our data, we ask "Why?" and it is at this point that the whole problem of philosophy of science arises. Why is the sky blue? Our answers seem always of the form "Because of such and such." We are all familiar with the inquisitive child who asks, "But why such and such?" and on being given the answer "Because of such and such" asks "Why because of such and such" and the parent soon loses his kindliness and says, "Don't bother me." The child is, of course, correct in his questioning. And the parent can only answer at the end by a statement of the sort, "That is the way things are."

Consider the "simple" question of the adult male of 21 who is uncomfortable with being 80 inches tall and asks "Why am I 80 inches tall?" To make the obvious analogue in the mental arena, consider the question "Why is my IQ 80?" And to give the question the obvious relevance, note that this individual with IQ 80 will not be able to function reasonably as, say, a bank teller, or in the very great majority of human occupations of our present society.

We surely wish to be able to answer either of the above "simple" questions. What, in fact, is the usual nature of an answer, "Because of such and such"? It seems to be of the form (i) At some previous times you had the attributes such and such, and (ii) It is our *opinion* that such and such attributes at times $\{t\}$ *lead to* the attribute, "height is 80 inches at age 21". Of these two assertions, presumbably (i) is unquestionable and the problem lies in (ii), in which the reader will note that I have emphasized the words "opinion" and "lead to". What entitles one to have the opinion given in (ii)? There are, obviously, two routes and it is critical to differentiate between them. I call these the statistical route and the experimental route.

The statistical route is simply that one has a large number of instances of (x, y), where x = height at 5 years and y = height at 21 years and data analysis of a very simple sort tells us that in that historical set of data, individuals who were around 60 inches at 5 years were around 80 inches at 21 years.

The experimental route is in strong contrast, though there seems to be an appalling ignorance as to what constitutes an experiment as opposed to an observational study. Suppose that we have at our disposal intervention processes that will make the height of 5-year olds to be 40, 45, 50, 55 and 60 and that we then observe individuals, whose 5-year height has been caused by intervention, at the age of 21. Suppose those whose height at 5 was 50 inches were around 72 inches at age 21, while those whose height at 5 was 60 inches, were around 80 inches at age 21. If we could have an experiment, well replicated, of course, which gave this result we should, I submit, be entitled to take the view that the intervention that caused the height to be 60 at 5 years caused the height to be 80 at 21 years.

With this experimental route we may reasonably take the view that the height at 21 years was caused by the intervention—it was "due to" the intervention. In the former case, however, the use of the words "due to" is entirely unwarranted. I am sure I have sinned with essentially all other statisticians in this verbal misuse.

A reaction might be "Who cares?" It is just a matter of word usage. The words mean "What I want them to mean to me." But, in fact, much of the IQ controversy is based on a total semantic error; the data can only give us "due to" in the statistical (and incorrect) sense, and a single worker having satisfied himself on "due to" in the statistical sense, in the same writing and almost in the same breath uses "due to" in the intervention sense. The story is an old one with the hoary joke that the way to reduce the divorce rate in England was to reduce the imports of bananas.

5. *What is a Viable Notion of Causation?*

It may well be regarded as arrogant for me, a mere statistician, to raise this question. The history of the topic extends over centuries and even millennia; even to list individuals who are

regarded as founders and leaders of philosophy could occupy many pages and a discussion of what they say could easily occupy a heavy tome. To compress this history into a few pages is hopelessly difficult. But it is necessary for my exposition to give some of the portions which I regard as significant. Where should I start? A readable view on the matter is given by Nowell-Smith (1960) in *Encyclopedia Britannica.* The word 'cause' has been used in several senses: Sense I, Human Agency; Sense II, Causes in Nature; Sense III, Causes as Explanation. In Sense I, "to cause an event to occur is to perform an action with the expectation and intention that the event will follow". It is clear then that a cause must precede an effect in time, and that one cannot assert that A causes B unless A is an action and B is an outcome, and one has accumulated evidence that action A produces outcome B. One is surely involved then in comparative experiments. In Sense II, "a cause was *thought to be,* not merely the sufficient condition of its effect, but something which had the power to produce it" (my italicizing). Nowell-Smith then says: "To discover the cause of something is to discover what has to be altered by our activity to produce or to prevent that thing." But there seem to me to be problems in both these remarks. Note well the use of "the sufficient condition". We have to look at the three words: "the" implying one; "sufficient" meaning one does not know what; "condition" meaning one does not know what, again. Also, with regarded to be "what has to be altered", we are essentially back to Sense I, but we must then ask what we are to do about establishing "What has to be altered?" Finally, as regards "Cause as Explanation" I have already commented, and I should add my view that the use of cause in the sense of explanation is more often than not explanation in terms of something else that is deemed not to need explanation or is based on a hypothetical (*gedanken*) experiment, such as "nailing down" electric charges and measuring forces. Often the use of cause in "cause as explanation" is no more than sophistry (the sort of argumentation used against sophists by the greatest sophist of all time, Socrates).

The use of the phrase "what has to be altered" immediately poses the question of what we do and how we talk about an outcome which depends on the interaction of two forces, by which I mean two entirely distinct aspects which may be modified by human activity? The point is relevant in the whole IQ controversy, I believe. I shall go into this more later, but we clearly have two "forces", genome and environment. We can certainly imagine altering environment, e.g., we feed babies vitamins. We are not at the point yet of altering genome, though there is already much discussion and controversy about genetic engineering.

Let us suppose that we are able to replace the PKU gene (P) by a normal gene (p) in a fertilized human egg. Adjoin this to the already demonstrated fact that we can restrict the nutritional intake of phenylalanine. Then we construct a 2×2 table:

	Environment	
Gene	1	2
p	Normal	Normal
P	Normal	Idiot

Here environment 1 is free of phenylalanine while environment 2 has a "usual" amount. I pass over as not needing comment the attributes "normal" and "idiot". Now comes the question. What is the cause of the outcome "idiot"? Obviously, under my assumptions, the variation p or P and the variation environment 1 or 2 can be made. We have a clear case of joint causality by two "forces". Can we partition the causality in *any* meaningful way? I believe not. I shall return to this later.

Some more extended discussion of causality is relevant and bears on the nature-nurture question as well as general uses of ideas of causation. It appears (Stebbing 1931, p. 282 of Harper Torchbook reprint, 1961) that Bertrand Russell was in favor of the view "that causal regularities are nothing but observed regularities of sequence". Stebbing also says, "If Mr. Russell's view be correct, then every regular sequence is causal, since there is nothing more in the notion of cause than regularity of sequence." Russell talked about hooters. "For example, men leave a factory for dinner when the hooter sounds at twelve o'clock. You may say the hooter is *the* cause of their leaving. But innumerable other hooters in other factories which also always sound at twelve o'clock, have just as much right to be called a cause." I have to say that I, a lowly statistician, am compelled to regard Russell as being very stupid in this connection, for the reason that he did not, it would seem, give the slightest recognition of the idea of experimentation. The first student in statistics would merely suggest that one might one day not sound the hooter at that one factory until 12:05 pm and then note that the hooters of other factories did not produce leaving at 12:00 noon.

The incredible aspect of the whole business for me is that top minds of the world have written at length on causation and have written material which is useless, at best. It would seem that the notion of experimentally varying forces and observing the results, which is the basis of all good science, is given essentially no place.

In this connection the history of astronomy and mechanics is interesting and, I believe, informative. One had the Copernican theory or model. Then Kepler analyzed the data of Tycho Brahe and formulated his three laws. I believe it must be asserted that Kepler's laws were mere observational regularities. It was only when Galileo did what we term these days as experiments, or comparative experiments, that a start could be made towards the development of a causal theory. It is all too easy to forget the experimental origin of Newtonian mechanics and to view the whole game as a pure mathematical theory which, because mathematics was clear and perfectly logical, could not be challenged as the nature of the real world. The naiveté exhibited 'all over the shop' in this era of, say, 250 years ago must surely be regarded at the present time as incredible. If everything can be reduced to a correct mathematical argument, it must necessarily be *the* correct description of the real world. The argument breaks down in one way that was surprising (but perhaps should not have been surprising) with regard to the question of whether or not a mathematical argument can be proved to be correct. We have not got around yet to telling our students in basic mathematics that we cannot prove correctness. All we can say, it seems, is that if we accept such and such processes as constituting a correct mathematical argument, then such and such a proof is a correct mathematical argument. One can only comment "How are the mighty fallen!" The other way in which the argument breaks down is rarely discussed, and the only readily available account I know of is that of Northrop (1948). We may construct a theory by mathematical processes as exemplified by the following simple example: we have two variables y and x which, we suppose, can vary over, say, intervals of the real line and we know that $d^2y/dx^2 = -k^2y$, (with all the undescribed substructure here involved). Then we can deduce, if our mathematical processes are accepted as perfect, that $y = \alpha \sin(kx + \beta)$. But this is just a string of WFF's (well-formed formulae). We have manipulated ideas, or to give the affair an interpersonal existence, we have written on a piece of paper a string of symbols. What can this string tell us about the existent real world? Nothing surely, unless we can establish a correspondence between the symbols (e.g., written on the piece of paper, so they exist objectively) and measurable attributes of space-time. How do we establish epistemic correlations and how do we validate such epistemic correlations? It is obvious surely, that we cannot validate these perfectly.

An answer to this dilemma—or rather a suggestion, because there can be no perfect

answer—is that of Peirce. He said somewhere, "Any statement that claims to be true must have practical consequences" (presumably verifiable) and "Truth is what you can do with statements." To this should be added the following sort of statements:

(i) No theory is true,
(ii) No theory can be claimed to be partially true unless it has been subjected to experimental test.

By experimental test, I have in mind the following sorts of things:

(a) If you alter the environment from birth, you will not alter IQ at age 8.
(b) If you alter the environment from lac-minus to lac-plus, you will find that bacteria of this strain will grow at normal rates.
(c) If you transfer (by some yet unknown procedures) 25% of the genes of this Caucasian population into that African population, at gamete formation or at birth, keeping everything else the same, *you will find that IQ has been raised, say 10 points.*
(d) To take a possible sociobiological example, if you transfer 25% of the genes of race A of ants into race B of ants, keeping everything else constant, you will find that a standardized measure of, say, dancing propensity is increased by 10 units.

In these examples, the names are really quite irrelevant. Perhaps ants do not dance. If so, pick some other quantifiable behavioral attribute. Part of my intended message in the present paper is that these statements should lie at the core of all scientific activity, and that neglect of them is unfortunately widely diffuse in areas that term themselves branches of science and, most unfortunately, in the human IQ arena. Even if the data and the data analysis in the human IQ arena are correct, what entitles one to infer causality in the explicit and definite sense of intervention results that I have given?

6. *Data Analysis and Analysis of Variance*

The whole area of the heredity and environment controversy is replete with data analysis. Just as the epistemology of the idea of causation needed discussion above, the nature of data analysis in general, and in the present arena, must be discussed and in this, as before, it is necessary to go down to basics. A young person now entering the field of statistics might well be presented with the view that a new sub-discipline of statistics has been created in the past decade or so, this having the name "data analysis". The origin and epistemology of this activity must be discussed.

Let us begin by stating some "obviosities". Data analyses aimed at discovering uniformities of nature have been used since time began. If we take the view of an individual who processes sense perceptions in such a way as to alter a mode of living, so as to avoid a particular unpleasantness, then we have to conclude that data analysis permeates biological life. Simply consider the homely examples of cats and dogs. The data analyses of humans are very sophisticated. But surely the data analyses of animals are very sophisticated—so sophisticated, often, that we have no understanding at all of how they do them. The distinguishing feature of *Homo sapiens* is that he makes a semi-permanent record of his data, his data analysis processes and the results.

What then is data analysis? What is the epistemology and logic of the activity? To discuss this at appropriate length is impossible here. But the nature of correlation, regression, and analysis of variance must be discussed. To make these matters clear, I deem it entirely appropriate to talk about populations.

Mathematical statistics is dominated by sampling theory, but we may ignore a host of very difficult mathematical problems, and we may attack the basic epistemological questions without becoming involved in these. I shall attempt to present and sustain the view that data analysis is essential but is *only* a first step in scientific exploration. I believe the topic needs discussion because society is being flooded by data analyses on the IQ controversy. We may, I believe, correctly understand aspects of the procedures *only* by considering that we have a population of individuals on which we have observations. We suppose that we have a measurement process and that our individuals are named or labelled, so that in the simplest case we have numbers $x_1, x_2, \cdots, x_N$. What can we do with this array of numbers? Obviously, we can make a histogram; if we read J. W. Tukey (1977), we shall construct stem and leaf diagrams, hanging rootgrams and so on. Why should we do this? We may transform our x's in a zillion ways and repeat these operations with this derived array of numbers. We can keep a modern computer and graphing device busy for hours or days. But why? What are we trying to achieve? Perhaps it is appropriate that I give my opinion that the devotees of data analysis do not tell us what they are trying to achieve. They tell us to do this and that, but they have not told us what their aims are. I will do some of these actions also and when I do, I am trying to find structure in the data. What do I mean by this? I want to see if my data have the appearance of one or other configurations I have met in the past or which probability modelling suggests, e.g., a mixture of two Gaussian distributions or a Laplace distribution, or whatever.

Suppose I do, in fact, find that my population of numbers is like a mixture of two Gaussian distributions. Am I then through? Do I merely report this impression of the data to an anxiously waiting world and turn to another problem area? I suggest that the answer is negative. I have really accomplished very little. My idea of a mixture of two normal distributions must be the start of real investigation. If the individuals are cats and my x is a measure of activity in a particular testing situation, I should contemplate other testing situations and get new data on the individuals in a new situation, and determine whether the partition "stands up". Or if the individuals are plants, and I can obtain progeny, I should obtain, perhaps, an offspring from each and apply the measurement process to the offspring. Now all this is rather obvious. It has been obvious, I surmise, for centuries. But it seems to receive very little attention. I am rather inclined to the view that a single data analysis, by itself, of itself, contributes very little to the accumulation of knowledge. If, of course, the data analysis is seen by some other investigator who is then stimulated to make similar data analyses in situations like the one analyzed, there is some possibility of fruit. Data analyses give suggestions, and one hopes useful suggestions, but if the suggestions are not followed up (or cannot be followed up) the activity must surely be regarded as rather useless.

Now let us suppose that on each individual we have two measurements. If both measurements are categorical we shall look at a two-way frequency table. But then so what? The data analysis must lead to suggestions on other potential sets of data that may be enlightening, *and* some other data sets must be obtained, or the exercise is rather empty, like banana imports and divorce rates. The contribution of statistics in the past hundred years, and, I believe, a remarkably fine one, is the development by one route or another, but usually involving the probability calculus, of various modes of looking at such a data set. The most popular one is to apply the analysis of variance that is applicable to this case.

The analysis of variance is 'simple' for one data configuration, the so-called one-way classification. Here we have observations labelled by y_{ij} in which individuals are classified by one factor with levels indexed by i and individuals are indexed by (ij). The obvious identity is

$$y_{ij} = y.. + (y_{i.} - y..) + (y_{ij} - y_{i.})$$

where $\{y_{i\cdot}\}$ denotes averages, and this leads to

$$\sum y_{ij}^2 = \sum y_{i\cdot}^2 + \sum (y_{i\cdot} - y_{\cdot\cdot})^2 + \sum (y_{ij} - y_{i\cdot})^2$$

and this is the analysis of variance. There can, obviously, be no implications from this identity as to the causation of $\{y_{ij}\}$. We may and do state that the total variability is $\sum(y_{ij}^2 - y_{\cdot\cdot}^2)$, and that the portion $\sum(y_{i\cdot} - y_{\cdot\cdot})^2$ is associated with the factor of classification. There is nothing wrong with this statement. It is not a statement based on any assumptions. It is strictly one aspect of the array of numbers $\{y_{ij}\}$.

One can find a variety of presentations of analysis of variance which miss this elemental fact. In the heredity-IQ controversy, we see a statement by Feldman and Lewontin (1975) on the nature of analysis of variance that I believe to be totally wrong: "The analysis of variance is, in fact, what is known in mathematics as local perturbation analysis." In fact, perturbation analysis is a well-defined mathematical-statistical area, and the intersection of that area with the basic idea of the analysis of variance is essentially negligible. They also say "An analysis would require that we know the first partial derivatives of the unknown function $f(G, E)$." This illustrates a basic epistemological error. Suppose G and E were real variables and we did know the first partial derivatives; then, so what? They also say "The analysis of variance produces results that are applicable only to small perturbations around the current mean"; again a basic epistemological error. On the matter of the role of variance, to say that additive genetic variance is important "since Fisher's fundamental theorem of natural selection predicts . . ." is wide of the mark, and again exemplifies an error commonly made in population genetics. Fisher's theorem, *if it is correct,* deals with fitness, whatever that is (and population geneticists are curiously silent on the matter, using a symbol such as s, and rarely, if ever, discussing the matter of its epistemic correlation to some observation protocol; see, for instance, Kempthorne and Pollak (1970)). It is necessary to my general thesis to bring these matters into the discourse, because understanding of what the analysis of variance does and what it does not do is absolutely critical in the heredity-IQ controversy. These criticisms must *not* be interpreted to suggest that *all* of the Feldman-Lewontin paper is suspect.

We now turn to what I regard as a shocking error of logic. Suppose

$$\sum(y_{i\cdot} - y_{\cdot\cdot})^2 / \sum(y_{ij} - y_{\cdot\cdot})^2 = 0.80,$$

so that 80% of the variability is associated with groups. It is all too easy to go from this to the totally erroneous statement that 80% of the variability is *due to* the factor of classification, where *due to* is interpreted as *caused by*. What has gone on in the IQ-heredity controversy is little more than this. It is obvious that this analysis of variance can tell us *nothing* about causation.

The next step in the ideation is the matter of a population of numbers with *two* crossed factors, individuals of which we can represent by y_{ijk}. Here i could label genotypes and j environments. What now is the analogue of the simple analysis of variance in the previous case? If we represent the averaging over a subscript by replacing it with a dot, we have, *obviously,*

$$y_{ijk} = y_{\cdots} + (y_{i\cdot\cdot} - y_{\cdots}) + (y_{\cdot j\cdot} - y_{\cdots}) + (y_{ij\cdot} - y_{i\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdots}) + (y_{ijk} - y_{ij\cdot}).$$

We may then follow the obvious path of squaring and adding. We have then (because the square of a sum is the sum of squares plus the sum of crossproducts),

$$\sum y_{ijk}^2 = \sum y_{\cdot\cdot\cdot}^2 + \sum(y_{i\cdot\cdot} - y_{\cdots})^2 + \sum(y_{\cdot j\cdot} - y_{\cdots})^2 + \sum(y_{ij\cdot} - y_{i\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdots})^2 + \sum(y_{ijk} - y_{ij\cdot})^2 + \text{Remainder},$$

where $\sum$ denotes summation over i, j and k and "Remainder" is equal to a sum of sums of crossproducts. We must obtain a non-shallow idea of what is happening here.

Suppose the frequencies in our population are proportional, so that if n_{ij} is the number of individuals in the (ij) cell, it is the case that $n_{ij} \propto r_i s_j$. Then it is elementary algebra to see that the sum of any crossproduct terms of a particular type is zero. It is not at all so obvious to see that if the frequencies are not proportional, the "remainder" will not be equal to zero for an arbitrary collection of numbers $\{y_{ijk}\}$. I shall not bother to give a proof (which should be a standardly presented matter in a reasonable course on analysis of variance) that the remainder is zero only if the frequencies are proportional.

Let me suppose that the frequencies are proportional. Then we shall say that the portion $\sum(y_{i..} - y...)^2$ of the total variability is *associated* with the factor with levels represented by i, and that the portion $\sum(y_{.j.} - y...)^2$ is *associated* with the second factor with levels j. I take up the non-proportional frequencies case in Section 9.

It is then no step at all, except that it is completely unjustified, to say that the portion $\sum(y_{i..} - y...)^2$ is *due to* the i-factor, and the portion $\sum(y_{.j.} - y...)^2$ is *due to* the j-factor. This is done again and again. With only minor (logically) additions, this is what the calculations of heritability of IQ do. Obviously, I cannot here give the full story, but I assert with complete confidence and with an open challenge to be refuted that this is the basic 'logic' (or lack of logic, to be more precise) that underlies the whole game.

Let us go a little further and suppose that we have a very large data set, i.e., a very large population, which satisfies the condition of proportional frequencies. Let us suppose (with average $= 1/N\, \iota_i \sum_j \sum_k$), we find that average $(y_{ijk} - y...)^2 = 225$, average $(y_{i..} - y...)^2 = 180$, average $(y_{.j.} - y...)^2 = 45$, average $(y_{ij.} - y_{i..} - y_{.j.} - y...)^2 =$ zero essentially, and average $(y_{ijk} - y_{ij.})^2 =$ zero, essentially. Then it is said 80% of the variability is *due to* heredity and 20% of the variability is *due to* "random" environment.

The next steps need a little exposition. Let us suppose that we are able to represent the genomes and environments on a linear scale and that when this is done the effects are normally distributed. Then something of the following sort is said:

> *Obviously a variance of 45 is due to environment. So if we could improve by environmental intervention the environmental conditions by two standard deviations, we would increase IQ by only* $2\sqrt{45}$ *or say, 14 points. Group 1 of individuals is below group 2 of individuals on average by 14 points. So giving group 1 individuals a two-standard deviation improvement in environment (which means with normality raising to the 95 percentile of environmental favorability), we would bring the average of group 1 up to that of group 2.*
>
> *On turning the argument around, to "explain" the 14-point difference we have to hypothesize that the environment of group 1 is at the 5 percentile. But to hypothesize this is ludicrous. Therefore group 1 is lower because of factor 1—heredity.*

Then one goes further and says "Obviously, intervention can do very little". Also, if two large groups differ in mean, that difference cannot be due to environment. I am not deliberately attempting to make a caricature of a portion of the literature, although the reader might suspect so. A referee has asked for actual references here. I suggest that the reader pick a work *at random* from the writers who advocate "heritability" of, say 80%. Jensen (1969) is one source.

Let us suppose that we can in fact classify our large population by the two factors (we cannot, of course), and suppose we found proportional frequencies and suppose our data analysis gave the above results. Then we must ask: Does this analysis tell us anything about the causation? Does the analysis tell us anything about the potential effects of intervention?

The answer must clearly be in the negative. This sort of analysis cannot by the greatest stretch of the imagination tell us about causation or how much we can "boost" IQ by intervention. Even under the most favorable circumstances, a data analysis of an existent population cannot tell us about the potential effects of intervention unless that existent population has sub-populations which differ only in that intervention has been exercised by the observer, i.e., in a comparative experiment.

The above description of simple cases of analysis of variance clearly brings to mind a dilemma. Take the case of the one-way classification with the array $\{y_{ij}\}$. Now surely we can take $z_{ij} = f(y_{ij})$, where the relationship is 1 − 1. Then we can do our analysis of variance on $\{z_{ij}\}$. Of course we can! And if we do, we shall then get different numbers and if we use the common language, we shall say, for example, that, say, 60% of the variance of $\{z_{ij}\}$ is *due to* the factor of classification. If, now, this factor is genome, we shall now say the heritability is 60%, and then, with no trouble at all, we shall have the result that 60% of the attribute, say IQ, is *caused by* genome. Here I do not bring in the fact that there will be interaction with respect to z. There is "nothing wrong" with the "60%", because it refers to z.

Let us continue by considering a different data structure in which we have a *population* of 2 vectors (x_i, y_i), $i = 1, 2, \cdots, N$. Then it is no trick at all to consider approximate description of the vector $y' = (y_1, y_2, \cdots, y_N)$ by the vector $x' = (x_1, x_2, \cdots, x_N)$. To make a correspondence with the genetic situation, suppose in our population we have (x_u, y_u), $u = 1, 2, \cdots, k$, with frequency of the u*th* vector equal to w_u. Then we merely find α and β to minimize

$$\sum_u w_u(y_u - \beta_0 - \beta_1 x_u)^2.$$

This trivial task gives us the well-known answers and finally the *numerical* identity

$$\sum w_u(y_u - \bar{y})^2 = \beta_1^2 \sum w_u(x_u - \bar{x})^2 + \text{Remainder}.$$

Suppose the L.H.S. is 12000 and the first term on the R.H.S. is 9600. Then we say 80% of the variance in y is *due to* variation in x. So, for instance, we might say that 90% of the yearly variation in birth rate is *due to* quantity of bananas imported. And why not? If we use language properly and restrict the phrase "due to" to causal, i.e., comparative experimental, contexts, we should say 80% of the variability in y is associated linearly with variation in x. And when we say this, we are not confusing the world and our readers, *and ourselves.*

It is elementary and amusing to apply this to a simple case of 2 dichotomous variables. Suppose we have the population configuration

x	y	relative frequency
0	0	p
0	1	q
1	0	r
1	1	s

Then our little analysis of variance is as follows (using the erroneous language):

Due to	Mean Square
x	$(ps - qr)^2/(p + q)(r + s)$
Residual	$pq/(p + q) + rs/(r + s)$
Total	$(p + r)(q + s)$

The real scientist, *given only this background* and this analysis of variance, should react with the question "So what?" and a comment such as, "Let's watch a big tennis game on TV."

7. Correlation, Causation and Path Analysis

I find it highly ironic that this topic was first addressed *seriously* only around 1920 by Sewall Wright (1921). I shall not discuss the general nature of the development view of various presentations (e.g., Kempthorne 1957, Chapter 14, Li (1975), and *of course* many writings by Wright himself which are cited in both these works). However, a few comments seem both appropriate and necessary.

I believe it fair to say that Wright's development consisted of two important components:

(a) developing consequences of an assumed stochastic process, namely Mendelism,

and

(b) developing rationales for the analysis of observational, i.e., nonexperimental, data.

I find it interesting that the contrast between observational and interventional studies was not discussed by Wright. I believe that the absence of this is a severe defect. With respect to component (*a*) above, there can be no question of the meaning of the theory; there cannot in general be any question of the meaning of any "pure" theory other than the question of whether propositions that are supposedly derived are, in fact, proved within the usually implicit rules of proof. One does not ask what the meaning is of the theory of groups or of complex variables. To use an existential phrase: "The theory is". Whether Wright's ideas are good for Mendelian calculations is not a question relevant to the present discourse, so I shall not discuss it.

I must, however, discuss the second component. Wright does not, I believe, discuss any epistemological aspects of his development, e.g., what he means by causation. He tells us very important ideas, which I will enumerate, briefly, but I hope in a form which he would accept, more or less:

(a) Suppose we have a population *of t-vectors* $(x_{i1}, x_{i2}, \cdots, x_{it})$, $i = 1, 2, \cdots, N$. *(N may be infinite).*

(b) Then we may, obviously, condense the 2-way table [*rows (= individuals)* $\times$ *columns (= traits)*] *by the column means, by the column mean square variabilities, e.g.,* v_{ii}, $i = 1, 2, \cdots, t$ *and by the covariabilities of different columns,* $v_{ii'}$, $i \neq i'$.

(c) We may then calculate the population mean square linear regression of any trait on one or more of the other traits.

(d) The resulting regression equations may be used to predict for an arbitrary individual in the given *population the value of any trait from values for that individual for other traits included in the regression (whether this prediction will be good or poor is a side issue).*

(e) Now, using Wright's (1921) words, this does not provide a means of "measuring the direct influence" of a factor that varies in the population. The observed regression or correlation gives "merely the resultant of all connecting paths of influence".

(f) So "it would often be desirable to use a method of analysis by which the knowledge that we have in regard to causal relations may be combined with the knowledge of the degree of relationship furnished by the coefficients of correlation".

The suggested method of doing this is the method of path coefficients, or what is nowadays termed path analysis. The upshot of the method is that if the correlations (or regressions) determined from the data are combined with a path diagram which is said to exhibit the assumed causal structure, one can then determine the strength of various directed links *in the assumed causal structure.* Furthermore, without injection of this mode of data interpretation, one may indeed obtain trash as regards causation (as in the divorce rate-banana example). But before passing on, I must emphasize that what is obtained is *not* necessarily trash when used for prediction. This general idea is very important and I regard Wright's work as a very significant contribution to ideas of data analysis and interpretation.

But, having said this, I must adjoin what I believe to be valid criticisms. My early example above of height at age 21 is, I believe, highly relevant. In an observational context, it is supposed (by the users of path analysis, including Wright himself) that we can take the view exemplified by the following proposition: Height at 5 years of age determines, at least partially, height at age 21. I believe I have to take the viewpoint that this is a "garbage" statement, *as it stands.* It may be that if we could alter height at 5 years, we would find height at 21 altered. But we might not find this and the only way we can determine whether this is the case is by some sort of intervention. I suggest that it is epistemological nonsense to talk about one trait of an individual *causing* or determining another trait of the individual. I regret that I have to take the view that a very fine intellectual creation, path analysis, has led to a flood of work purporting to establish or suggest strongly that certain attributes *cause* other attributes.

8. Independence

This is a word that should be banned from the literature, or we should agree to attach subscripts for various meanings. We have independence of random variables as a well-defined concept. In our two-way table, we considered the case of proportional frequencies. By very unfortunate (though 'natural' and perhaps inevitable) misuse of language, it is stated that if we have a two-way table with proportional frequencies, then we say that the row factor and the column factor are independent. I suggest that this is not merely unfortunate but downright erroneous.

If we think of a deity giving at *random* to a soul a collection of chromosomes and an environmental mixture, and that this deity makes a random draw from a population of chromosomes and a random draw from a population of environments, then we can, I suppose, agree that for that deity and for the souls, genotype and environment are independent. But we are not the deity.

What should it mean to say that genotypes and environment are independent? Only, I suggest, that there is a deity or some collection of forces that is making random drawings. And this is patently false except under a hypothesis of predetermination that must surely be regarded as ludicrous.

I will use independence$_2$ to denote the occurrence of proportional frequencies in our 2-way table of genotypes and environments.

9. Consequences of Falsity of Independence$_2$

We can see that if we have independence$_2$, then we can construct an analysis of variance of simple form. Let us use for *aata analysis purposes* a linear approximative model

$$y = \vartheta\mu + X_a a + X_b b + \text{residual},$$

where I shall not bother to explain standard notation except to say that $\mathcal{G}$ is a column of unities and X_a, X_b are incidence matrices.

If we have proportional frequencies, we have a single analysis of variance, which I may characterize completely by the "keying-out" of sources:

$$\mathcal{G}$$
$$X_a \mid \mathcal{G}$$
$$X_b \mid \mathcal{G}$$
$$\text{Residual.}$$

It is almost inevitable then to say that the sum of squares for $X_a \mid \mathcal{G}$ is *due to* factor a and the sum of squares for $X_b \mid \mathcal{G}$ is *due to* factor b. I have given my view that this is incorrect language, and I do not need to flog this topic any more.

But now suppose we do not have proportional frequencies. Then there are two analyses of variance that are relevant to the *supposed* causation:

$\mathcal{G}$		$\mathcal{G}$
$X_a \mid \mathcal{G}$	and	$X_b \mid \mathcal{G}$
$X_b \mid \mathcal{G}, X_a$		$X_a \mid \mathcal{G}, X_b$
Residual		Residual

If now one wants to use sums of squares (or mean squares, this being a trivial modification which I pass over), then one has problems—deep problems. What is the sum of squares "due to factor a"? Is it $SS(X_a \mid \mathcal{G})$ or $SS(X_a \mid \mathcal{G}, X_b)$?

This matter has been lying around the field of statistics since the 30's and the simple answer, relevant to some discussions, is that we should look at $SS(X_a \mid \mathcal{G}, X_b)$. Then if we want to look at the sum of squares "due to" factor b, we should look at $SS(X_b \mid \mathcal{G}, X_a)$.

Some discussion of this was given by Kempthorne (1957, Section 14.13). The idea given there was taken up by several workers and the term "analysis of commonality", or some such, has been brought in. I shall not discuss the problem here and shall merely cite Emigh (1977) who discusses at length the difficulties in the context of partitioning variance "due to" heredity and environment.

10. The Case of Random Variables

Suppose we have random variables Y, X_1, X_2, X_3, such that $Y = X_1 + X_2 + X_3$ and, to give a correspondence to some heredity literature, suppose X_3 is independent (a mathematically correct usage) of X_1 and X_2. Then we have with obvious notation

$$\sigma_Y^2 = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2 + \sigma_3^2.$$

If we use our unfortunate language, we say that a portion σ_3^2 of σ_Y^2 is "due to" X_3 and we then have to say how much is "due to" X_1 and "due to" X_2. This is a fertile plot for garbage again. I shall not make citations. What are we to say? For instance, can we say the amount σ_1^2 is "due to" X_1, or the amount $\sigma_1^2 + \rho\sigma_1\sigma_2$ is "due to" X_1? Either statement is unacceptable as Sewall Wright indicated (in less polemic language) around 1920.

It is interesting that the notion of "due to" was thought to be viable in the presence of independent random variables (i.e., $\rho = 0$). The point is really, I believe, that the notion of

"due to" is not viable in either case. I shall, however, later indicate senses in which partitions of variance are useful.

Before leaving the random variable case, suppose that we have two factors a and b, with b nested in a. Then, just using the identity

$$Y = \mathcal{E}(Y) + [\mathcal{E}(Y|X) - \mathcal{E}(Y)] + \{Y - \mathcal{E}(Y|X)\} = \mu + X_1 + X_2, \text{ say,}$$

and the obvious extensions, it is the case that $\mathrm{Cov}(X_1, X_2) = 0$, and we have no difficulties in naive phraseology that $\mathrm{var}(X_1)$ is "due to" the nesting factor and $\mathrm{var}(X_2)$ is "due to" the nested factor.

This, with extensions that are fairly straightforward but not obvious, is the basis for partition of variance between and within lines, and under appropriate circumstances we can make the partition as described, e.g., by Falconer (1960).

If now the effects of different genetic factors are additive and environment is a "random" additive factor, we can have a "simple" partition of variance.

But if the effects of genetic factors are not additive, this partition of variance cannot be made unless genotypes and environments are randomly associated. Emigh (1977) gives a discussion of this specific point. A basic error, of a statistico-logical type, has been made and received wide use by those who are commonly classed as "hereditarians".

11. Is Partition of Variance Sometimes Useful?

The analysis of variance and partition of variance has been widely used in plant breeding studies. My criticisms have no bearing on much of that activity. The studies are experimental. The "entries", e.g., lines, progeny groups, etc., are assigned to field plots at random, so within the context of the experiments, genotypes and environments are randomly associated. Also a strongly justified faith in the correctness of Mendelism "justifies" the use of associated ideas of partition of variance. This is not to assert that there are no residual problems. There are very many. One exposition of the role of analysis of variance *in quantitative genetics* is given by Kempthorne (1957), the ultimate impact being on choice of *planned* selection.

In the case of animal breeding, the level of unvalidated assumption is greater, but I shall not discuss this. Animal breeding is a curious mixture of observational and experimental science.

12. The "Logic" of Quantitative Genetic Theory

The reader may well question why in the above I have not introduced Mendelism. To answer is easy; almost all of the epistemology and logic of the IQ heredity controversy uses what I have discussed. Obviously, with Mendelian segregation, the same genotype occurs only rarely with so-called identical twins. The genetic approach is to conduct "controlled" mating experiments and observe categorizable phenotypes. This gives the classical qualitative genetics picture, so excellently given (and, if I may suggest, still unexcelled) by Sturtevant and Beadle (1940). With an uncategorizable phenotype one has to turn to quantitative genetics.

I shall confine myself to quantitative genetic theory in relation to an observational, i.e., not comparative experiment or interventional, arena. All we need to do is to read Fisher's (1918) title: "The correlation between relatives *on the supposition* of Mendelian inheritance" (my italicizing). This, the basic paper, I believe, is extremely difficult as I suggest above. It gives, as I note (Kempthorne 1957), an infinitesimal equilibrium theory of assortative mating. There are many obscurities over and above a most remarkable defect. This is that it is assumed that genotype and environment have additive "effects" (*not* effects due to causes,

cf. above), *and* that environment enters as a completely random force. We should not fault Fisher for this, and his paper is, I believe, a work of genius. But having said this I have to enter a caveat. While Fisher was concerned to show that observed correlations between relatives were explainable by his theory, and it was natural to have this aim at the time, the basic questions of validity of the model, particularly vis-a-vis environment, were not considered.

The logic of Wright (1921) is weaker in some respects and stronger in others. It is assumed *essentially* that a sperm and an egg carry "charges" which are then added, and then a random environmental effect is also added to give the phenotype. Again, I do not wish to denigrate a tremendous intellectual effort, which is of critical importance in humanity's tortuous path to achieve understanding. But it is high time to go beyond those first very, very difficult steps.

It is also important, I believe, to comment on the "biometrical genetics" approach emanating from Birmingham. I have commented elsewhere (Kempthorne 1977) on this in relation to applied quantitative genetics, e.g., in the development of better stocks in economically and foodwise important species. My opinion is that the theory has failed. Is then this theory likely to lead to advances in behavioral genetics or in the understanding of human mental abilities, as is often suggested? I believe not. I have to state, however, that I think this also was a great effort along the tortuous path to knowledge.

In connection with the Fisher and Wright models, it seems worth mentioning that it is a childish error to deduce from the facts (*i*) theory $\rightarrow$ "expected" results, and (*ii*) "expected" results $\doteq$ observed results that the theory is true. This is so often forgotten that one can only marvel. How often was it asserted in that "queen of sciences" physics that electrons and protons were *the* elementary particles?

The naiveté of the models is obvious. Even in the case of a "simple" plant like corn, the role of endosperm is not included. In the case of animals, we must surely entertain even at this naive level the possibility of maternal effects. These may be "under genetic control", *obviously,* but *also obviously,* there are purely environmental effects. The "theory" (and I deliberately use quotation marks here) does not take account of well-established maternal environmental forces and hazards.

What, indeed, is the "grip" on environment in the human IQ area? It is no more than "reared together" versus "reared apart", and what does "reared apart" mean? Nothing more than at some age two related individuals, e.g., identical twins or full-sibs were separated by adoption, and then placed in homes that could be related familially and/or, of similar economic and social nature. I can only comment: Really, how naive can one be? The Burt study was characterized in the literature as the "only experiment". Some experiment!

As soon as one turns to any behavioral measurements, the need to incorporate intrauterine, family and community environment is obvious. I have the view that the "hereditarians" are utterly naive. It is obvious that parental IQ influences offspring environment. It is obvious that there is cultural transmission. To ignore existence of this is merely stupid. I see no point in mincing words. If non-scientists *sometimes* have scorn for some supposed "scientific work", they should not be faulted.

To this naiveté in model formulation, must be added a statistical naiveté. Any statistical test has, within its conceptual underpinning, a sensitivity or power function. It needs essentially no deep thought to realize that the sensitivity of statistical tests for maternal effects, for genotype-environment interaction, etc., etc., is so low that to say, as has been said often, that such and such a model modification has been examined and found unnecessary is utter naiveté. And particularly so by virtue of "obviosities", such as the surely undeniable proposition that a Beethoven would not have come to light in an industrial slum, a city ghetto, or on a Mississippi slave plantation.

13. Heritability

There is a lot of discussion in the human IQ controversy about heritability. Jensen says it is of the order of 80%. Others do various data analyses and arrive at figures such as 40%. Kamin (1974) says (that he thinks) the heritability is zero. I arrive at my punchline. The argument is stupid from several points of view:

*(i) We will surely agree that in our population there is association of genome and environment (*though to nail this down is not easy*). So in view of what I have said above, we cannot reasonably state that such and such a proportion of variance is "due to" heredity. So "heritability" does not even exist in the human IQ context. Why then argue about the magnitude of an imaginary number?*

(ii) Even if in our data we had random association of genome and environment, we could perform our calculations without embarrassing problems, but we would still have a useless answer with respect to causation.

(iii) So the conclusions about the potential effects of intervention, e.g., a program such as Head Start, cannot be sustained from a logical viewpoint. We do not know from heritability studies "how much we boost IQ".

I suggest that Kamin's (1974) discussion, much of which I believe to be correct and well-aimed, can be interpreted to imply that he believes there is no genetic causation, a view which, from conversation, I know he does not hold, and a view which flies in the face of the spectrum of organismal knowledge.

14. The Burt Case

We have seen in the popular press a large amount of discussion of Burt's work. Did he "fake" data, for instance? It seems to be agreed by all protagonists that there are good grounds for considerable doubt on the data base he used. Two aspects need to be discussed. Even if Burt's data were "clean", the Burt studies are *not* experiments, they are merely uncontrolled observations, and the criticisms I give above are relevant. Furthermore, the fact that numerous other observational studies give similar correlations is of *no* relevance to the basic issue of causation.

A final irony, brought to my attention by T. H. Emigh, is that if we believe Burt and his nonsense about IQ being innate (a word *slopped* around by other notable figures), and then we define innate as caused by genome, then it is a tautology that IQ is 100 percent "heritable".

15. Data Analysis in the Area

The actual data analysis methods used in the arena are clearly of doubtful value. One exposition of the logic and defects was given by Emigh (1974). Goldberger (1977) has examined the writings and computations of many workers and demonstrates clearly, I believe, gross defects.

16. Can we boost IQ?

Jensen (1969) says, essentially, NO. He says, essentially, compensatory education has failed. I believe Jensen is *completely* wrong. I quote Palmer (1977), as merely one of many with the same view:

The evidence is that compensatory education has not failed. The only way to make a stab

at "How much can we boost IQ?" (phraseology that I find offensive) is to perform experiments—comparative experiments or intervention studies. Obviously, I suggest, the one experiment by Heber, Garber, Harrington, Hoffman and Falender (1972) is better than all of the data analyses of Burt, Jensen and others. See also Lazar, Hubbell, Murray, Rosche and Royce (1977) for the results of intervention studies.

17. Observational Science in General

I deem it appropriate to make some comments on the general problem which underlies observational science. We can go to Africa and observe a clan of monkeys; we can observe the behavior of ants; we can observe the geographical distribution of frequency of a particular gene; and so on. We can then develop in our minds a theory for what we observe, e.g., some teleological, mathematical, statistical, or stochastic model for our observations. To do so is really no trick at all, given the very fine array of expositions of various types of model and development of mathematical ideas. Then to advance hypotheses is again no trick at all. One can write a mighty tome, and one can "fill" science with small data sets and simplistic models. I shall not cite journals and books in which we see this, but I must mention IQ, meritocracy, educational inputs (e.g., the Coleman report) and sociobiology as currently interesting the public.

But something more must be done. There must be controlled experimentation. Let me give two examples:

(i) To find out if there is random mating, one must develop means of ascertaining the parentage of births.

(ii) To find out if a pattern of behavior is under the control of the nucleus, one must do experiments which attempt (but will never succeed completely) to eliminate the role of variations in environment (within a defined population of environments). So, for instance, to determine if a particular pattern of behavior is "caused by" genes and not by imprinting, maternal environment, and so on, one must experiment, e.g., by transplantation (in some species at least) of ova or by "creation" of zygotes in test tubes. And even then one has the problem of extrachromosomal inheritance. The resolution of matters depends critically on the development of techniques (e.g., use of radioactive traces, transplantation of ova, "transplantation" of genes, and so on). There can be no substitution for such techniques. Biometrical genetics of behavior traits is just a beginning, but some workers whom I shall not cite appear to think it is the end.

The moral, for me, is that observational science is the naive, but natural and very often the only possible, beginning. But to claim, pretend, or even suggest that this is the end is ridiculous. The second and absolutely essential stage is experimental or interventional observation. To many human scientists this will be no more than an obvious cliché. But ceaseless repetition is unfortunately necessary.

Behavioral science and sociobiology, to name two arenas, must move from observation to intervention or we should ignore them. To say this is very hard to do is no answer.

18. Prediction Versus Causation

I have alluded to this earlier. I regard the writings of Feldman and Lewontin (1975) as grossly defective on this. Let us take the prosaic problem of predicting the merit of the offspring of various bulls in a population of bulls. Then we are entitled to use any mode of prediction we can develop. The higher the degree of rationalism that is used the better, of course. But the "proof of the pudding is in the eating", no more and no less.

Now let us turn to the problem of genetic counselling. We wish to make a judgment of the possibility that a certain mating of humans will produce a particular type of defective. Then causation is irrelevant. I must search the available data to determine as good a predictive model as possible. If I do data analyses which suggest a threshold genetic model with additive gene effects on the underlying scale, well and good. I should use this model unless I have a better one.

To take the matter out of a genetic context, if data analyses tell me that in the available history individuals who smoke cigarettes heavily incurred lung cancer more frequently, then I would be foolish not to use this data analysis conclusion to predict which individuals will (or may) get lung cancer. If data analyses suggest that eating large amounts of cholesterol is associated with later clogging of the arteries, I should surely use the data analysis conclusion to warn the public (in spite of the potential economic loss to egg producers). This example is interesting because the egg producers say "You cannot prove that eating many eggs causes clogging of the arteries." The answer is simple, "Agreed; we cannot prove that. But our data strongly suggests a higher risk."

19. Concluding Remarks

I have attempted to advance reasons why the whole IQ-heredity argument as advanced by the hereditarians is deeply unjustifiable and strongly misleading. It has not been necessary to go beneath the overall surface of the reasoning that has been used.

Even if we use path analysis as a technique for *suggestions*, a path diagram that incorporates forces which we know to be present leads to impossibility of determination of path coefficients for *causal* inference. But if we ask if there is genetic causation of mental and behavioral traits in humans we have to keep an open mind. Extreme cases surely suggest so. In this respect, antihereditarians appear to be following emotions and desires (very laudable ones, to most of us) rather than reason. The anti-hereditarians, I suggest, destroy their case by taking illogical steps as well as ignoring "facts" they know well. On the former, for instance, it is a tautology that if there really are genotype-environment interactions, then there is genetic causation. By using erroneous reasoning this group has lent strength to the hereditarian "junky" reasoning.

Are there genetically caused differences between human races in mental and behavior traits? We do not know and it is incredibly difficult to nail this down. I have not discussed the writings of the Nobel Laureate, Schockley, on this matter. I believe them to exhibit the epistemological and logical errors I discuss in the paper, and hence to merit no attention. The omission from the references is deliberate.

We should not be surprised if we find, eventually, though I see no possibility of so doing in the indefinite future, that there are genetic differences just as there are differences in frequencies for some genes with manifest physical effects. We know that the human races have evolved under different environments which obviously require different mental abilities. There are, for instance, obvious differences in the hunting processes under which, say, Eskimos and Australian aborigines developed, and it is unreasonable to suppose that gene causation does not occur. The theory of evolution by natural selection has been so well substantiated that it cannot be dismissed, and in any case the theory of evolution by selection in animal and plant species is being validated all the time in our laboratory. All our biological knowledge and theory suggests that differences in genotype frequencies may have arisen.

To speculate on racial differences in mental and behavioral traits is not sinful but merely idle. The making of unjustifiable assertions on inferiority of some races is obviously and naturally insulting, and our black population, for instance, is entitled to be very angry.

To hope that the matter can be resolved by *observational* studies on soldiers in the Army or children we find in our society is highly naive. Even if there were no cultural transmission, without which we would have only sub-human creatures, the resolution would be difficult. We cannot nail down the heritability of height even. The existence of cultural transmission makes the resolution impossible, except by intervention experiments. The only intervention that is both available and acceptable to our society is environmental, and this cannot determine genetic causation.

Finally, I would not like the writing here to be considered as an argument against the use of statistical method and justified biometric approaches. Good tools can be used incompetently, and this has happened all too often in the present arena.

Acknowledgments

Journal Paper No. J-9022 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 1669. The research reported here was supported by the National Institutes of Health, Grant No. GM13827.

References

Burt, C. and Howard, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *British Journal of Statistical Psychology 9,* 95–131.

Emigh, T. H. (1974). Statistical methodology of the nature-nurture controversy in human intelligence. M.S. Thesis. Iowa State University.

Emigh, T. H. (1977). Partition of phenotypic variance under unknown dependent association of genotypes and environments. *Biometrics 33,* 505–514.

Falconer, D. S. (1960). *Introduction to Quantitative Genetics.* Ronald Press, New York.

Feldman, M. W. and Lewontin, R. C. (1975). The heritability hang-up. *Science 190,* 1163–1168.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh 52,* 399–433.

Galton, Francis (1869). *Hereditary Genius: An Inquiry into its Laws and Consequences.* Macmillan, London.

Goldberger, A. A. (1977). Models and methods in the IQ debate, Part I. Social Science Research Institute, University of Wisconsin, Madison, Wisconsin.

Heber, R., Gaber, H., Harrington, S., Hoffman, C. and Falender, C. (1972). *Rehabilitation of Families at Risk for Mental Retardation.* Rehabilitation Research and Training Center in Mental Retardation. Madison: University of Wisconsin, Madison, Wisconsin.

Jensen, A. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review 39,* 1–123.

Kamin, L. J. (1974). *The Science and Politics of IQ.* Halsted, New York.

Kempthorne, O. (1957). *An Introduction to Genetic Statistics.* Wiley, New York. (Reprinted by Iowa State University Press), Ames, Iowa.

Kempthorne, O. (1974). Review of collected papers of R. A. Fisher. *Social Biology 21,* 98–101.

Kempthorne, O. (1977). Status of quantitative genetic theory. *Proceedings of the International Conference on Quantitative Genetics.* Iowa State University Press, Ames, Iowa.

Kempthorne, O. and Pollak, E. (1970). Concepts of fitness in Mendelian populations. *Genetics 64,* 125–145.

Lazar, I., Hubbell, R., Murray, H., Rosche, M. and Royce, J. (1977). Preliminary findings of the developmental longitudinal study. Presented at Office of Child Development "Parents, Children and Continuity" Conference. El Paso, Texas.

Li, C. C. (1975). *Path Analysis—A Primer.* Boxwood Press, Pacific Grove, California.

Northrop, S. C. S. (1948). *The Logic of the Sciences and Humanities.* MacMillan, New York.

Nowell-Smith, P. H. (1960). Causality and Causation. *Encyclopedia Britannica.*

Palmer, S. H. (1977). The effects of early childhood interventions. Presented at the Annual Meeting of AAAS, Denver, Colorado.

Stebbing, L. S. (1931). *An Introduction to Modern Logic.* Methuen, London. Printed in *Harper Torchbook,* 1961).
Sturtevant, A. H. and Beadle, G. W. (1940). *An Introduction to Genetics.* Saunders, Philadelphia.
Tukey, J. W. (1977). *Exploratory data analysis.* Addison-Wesley, Reading, Massachusetts.
Vetta, A. (1976). Dominance variance in Fisher's model of assortative mating. *Annals of Human Genetics 39,* 447–453.
Weinberg, W. (1908). Uber Vererbungsgesetze beim Menschen. *Zeit. indukt. Abstamm. -u. VererbLehre 1,* 377–392, 443–460; *2,* 276–330.
Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20,* 557–585.

Received September 1977; *Revised November* 1977