Heritability Analyses of IQ Scores: Science or Numerology?

David Layzer

The question, "To what extent can the development of basic cognitive skills be influenced by various kinds of environmental intervention?" is central to current discussions of educational and social policy. Much of the discussion has focused on one rather narrow, but comparatively well-defined, aspect of this question—the heritability of scores on standardized tests and the implications of heritability estimates. In 1969, A. R. Jensen (*1*) reviewed a large number of British and American studies of the broad heritability of IQ scores and concluded that, for representative British and American populations, it probably lies between 0.7 and 0.8, the best single estimate being close to 0.8. Jensen argued from these figures that inequalities in cognitive performance are largely genetic in origin and that comparatively little can be done to reduce them through practicable educational and social reforms, a thesis he has since developed further (*2*). He also argued that the reported difference in average IQ between black and white children in the United States probably has a substantial genetic component, another thesis elaborated later (*3*). From the same heritability estimates, R. J. Herrnstein has argued (*4*, *5*) that the elimination of artificial barriers to social and economic mobility (such as those based on race, sex, income, and social class) must inevitably lead to the emergency of an hereditary meritocracy based on IQ.

After the publication of Jensen's 1969 article, Christopher Jencks undertook a detailed analysis of the available data on IQ. Devoting themselves to the broader question of how family and schooling affect social and economic, as well as cognitive, inequality, Jencks and his colleagues concluded (*6*, p. 315) that "the chances are about 2 out of 3 that [the heritability of IQ scores] is between 0.35 and 0.55." But Jencks' analysis also indicated that the purely environmental contribution to the IQ variance probably lies between 0.25 and 0.45—an estimate only mildly discordant with Jensen's estimate of 0.2 to 0.3. Jencks and his collaborators concluded, moreover, that educational inequalities accounted for only a minor part of the environmental variance.

From this brief summary, it is clear that two methodologically distinct kinds of issues are involved in current discussions of IQ heritability and its implications. Some of the issues are purely scientific: What are the limitations of conventional heritability analysis as applied to IQ scores? How reliable are heritability estimates like those quoted above? What inferences can legitimately be drawn from heritability estimates of various kinds? Of more widespread interest than these technical questions are those that involve social, political, educational, and philosophical considerations. Questions of the first kind can be discussed and resolved in a value-free context—or at least in a context of values agreed upon by members of the scientific community. Moreover, they must be resolved before the broader issues can be meaningfully debated. In this article I address myself to that task.

Although several illuminating discussions of IQ heritability have already appeared, it seems to me that some important aspects of the problem have not yet been adequately discussed. For example, nearly all the published discussions that I am aware of take it for granted that some meaningful estimate of IQ heritability—high or low, rough or accurate—can be extracted from the reams of published statistics and that refinements of current techniques for gathering and analyzing test data may be counted on to yield increasingly reliable estimates. These propositions are by no means self-evident, however, and one of my purposes here is to demonstrate that they are actually false.

This conclusion rests upon two arguments: One concerns the limitations of conventional heritability analysis, the other the validity of IQ scores as phenotypic measurements. Contrary to widely held beliefs, (i) heritability analysis does not require the genotype-environment interaction to be small, and (ii) a high phenotypic correlation between separated monozygotic twins does not, in general, imply that the genotype-environment interaction is small. If genotype-environment interaction does contribute substantially to the phenotypic value of a trait (as there are strong biological reasons for supposing in the case of phenotypically plastic traits), then a necessary and sufficient condition for the applicability of heritability analysis is the absence of genotype-environment correlation. This condition is rarely, if ever, met for behavioral traits in human populations. The second argument is that IQ scores contain uncontrollable, systematic errors of unknown magnitude.

Limitations of the Heritability Concept

Estimates of broad heritability (h^2) answer the question: What fraction of the variance of a phenotypic trait in a given population is caused by (or attributable to) genetic differences? A phenotypic value (P) that depends on a genotypic variable (or set of variables) (x) and an environmental variable (or set of variables) (y) may be expressed as the sum of a genotypic value [$G(x)$], an environmental value [$E(y)$], and a remainder [$R(x,y)$] that depends jointly on x and y (see Eq. 4). Under certain conditions (which probably never obtain in natural human populations), there is a well-defined optimal additive decomposition of this kind—that is, a decomposition that minimizes the variance (Var) of R. Under these conditions, the quantities $h^2 \equiv Var\{G\}/Var\{P\}$ and $e^2 = Var\{E\}/Var\{P\}$ (e^2 is the environmental fraction of phenotypic variance) are well defined.

Even when h^2 and e^2 are well defined, however, their practical significance may be obscure. Consider the hypothetical trait illustrated in Fig. 1. Note that the phenotypic value (P) cannot be expressed as the sum of a genotypic value

The author is professor of astronomy, Harvard University, Cambridge, Massachusetts 02138.

[$G(x)$] and an environmental value [$E(y)$]. Consider, for example, the pairs of points (A_1, A_2) and (B_1, B_2) in Fig. 1. Since ΔP (the difference between the phenotypic values) = 0 for each pair, the differences ΔG and ΔE would be equal and opposite if an additive representation were possible. But ΔG must have the same value for both pairs, and the values of ΔE must evidently be unequal [since they correspond to different increments (Δy)]. Hence an additive representation is impossible (*7*).

I show below that G and E can be unambiguously defined if, and only if, genotype-environment correlations are absent. Even then, however, a certain practical ambiguity persists. Genetic differences may influence the development of a trait in qualitatively distinct ways. For example, the curves labeled x_1, x_2, and x_3 in Fig. 1 have different thresholds, different slopes, and different final values. Heritability estimates do not take such qualitative distinctions into account. Thus, if the environmental variable y is distributed in a narrow range about the value y_1, as illustrated in Fig. 1, h^2 is close to unity. Yet in these circumstances the phenotypic variance could reasonably be considered to be largely environmental in origin since it is much greater than the phenotypic variance that would be measured in an environment ($y = y_2$) that permitted maximum development of the trait, consistent with genetic endowment. This point has been elaborated by R. C. Lewontin (*8*).

The conventional definitions of G (genotypic value) and E (environmental value) have specific mathematical advantages in the context of a specific mathematical theory. Other ways of assessing the effects of environment on phenotypically plastic traits may, however, be more useful in other contexts. In particular, certain kinds of intervention studies may provide more direct and more useful information about the effects of environment on IQ than conventional studies of IQ heritability.

Conventional Heritability Analysis and Its Limitations

Three mathematically and biologically distinct effects complicate conventional heritability analysis: genotype-environment correlation, genotype-environment interaction, and gene-gene interaction. I now consider the theoretical limitations imposed by these effects.

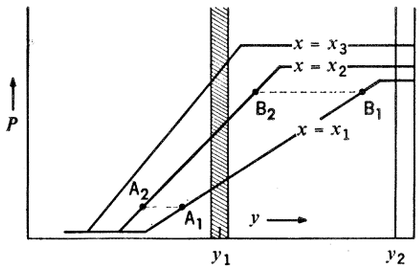

Fig. 1. Phenotypic value (P) of a hypothetical metric trait as a function of an environmental variable (x) for three values of a genotypic variable (y). A_1 and A_2 (also B_1 and B_2) indicate individuals with a common phenotypic value but distinct genotypes x_1 and x_2, respectively.

Let $P(x,y)$ denote the value of a phenotypic character that depends on n genotypic variables $x_1, \ldots, x_n$, denoted collectively by x, and m environmental variables $y_1, \ldots, y_m$, denoted collectively by y. The x_i and y_j (where i and j are integers) are random variables whose joint-frequency distribution is specified by the function $\Phi(x,y)$; P is a random function of these random variables. $\mathcal{E}\{f\}$ is the expectation value of a function $f(x,y)$:

$$\mathcal{E}\{f\} = \int \ldots \int f(x,y)\ \Phi(x,y)\,dx\,dy \qquad (1)$$

where $dx = dx_1 \ldots dx_n$, $dy = dy_1 \ldots dy_m$, and the integration extends over the ranges of all the variables.

The genotypic value G and the environmental value E are defined as follows (*9*):

$$G(x) = \int \ldots \int P(x,y)\ \Phi(x|y)\,dy \qquad (2a)$$

$$E(y) = \int \ldots \int P(x,y)\ \Phi(y|x)\,dx \qquad (2b)$$

where the conditional frequency functions $\Phi(x|y)$ and $\Phi(y|x)$ are defined by

$$\Phi(x,y) = \Phi(x|y)\ \Phi_2(y) = \Phi(y|x)\ \Phi_1(x) \qquad (3a)$$

$$\Phi_1(x) = \int \Phi(x,y)\,dy,\ \Phi_2(y) = \int \Phi(x,y)\,dx \qquad (3b)$$

One can then write P in the form

$$P(x,y) = G(x) + E(y) + R(x,y) \qquad (4)$$

(this equation defines R), whence

$$Var\{P\} = Var\{G\} + Var\{E\} + 2\,Cov\{G,E\} + I \qquad (5)$$

where

$$I = 2\,Cov\{G + E,R\} + Var\{R\} \qquad (6)$$

Here Cov denotes the covariance and I the contribution of genotype-environment interaction to the phenotypic variance.

More generally, consider two subpopulations between whose members a 1-to-1 correspondence, defined by some genetic or environmental relationship, has been established. For example, the two subpopulations might consist of children and their respective mothers or of foster children and their respective foster siblings. For each subpopulation, one may write an equation with the form of Eq. 4. Using variables without primes to refer to the first subpopulation and variables with primes to refer to the second, multiplying the two equations together, and averaging the resulting equation, one obtains

$$Cov\{P,P'\} = Cov\{G,G'\} + Cov\{E,E'\} + Cov\{G,E'\} + Cov\{G',E\} + J \qquad (7)$$

where

$$J = Cov\{G + E,R'\} + Cov\{G' + E',R\} + Cov\{R,R'\} \qquad (8)$$

where J represents the contribution of genotype-environment interaction to the phenotypic covariance.

The phenotypic variance, along with phenotypic covariances referring to certain pairs of subpopulations, may be estimated from appropriate phenotypic measurements. From these data one must try to estimate the individual components of the phenotypic variance that appear on the right side of Eq. 5. In what circumstances is this possible?

If either E or G is negligible, the problem has a trivial solution. Otherwise, it is insoluble unless the genotype-environment covariances in Eqs. 5 and 7 are negligible. For if these terms are not negligible and their structure is not known a priori, the number of unknowns exceeds the number of conditions. One cannot even derive an upper bound for the genetic variance under these circumstances. Suppose, to this end, one equates all genotype-environment correlations (Cor) in Eqs. 5 and 7 to their theoretical upper limits, $Cor\{G,E\} = Cor\{G',E'\} = 1$ and $Cor\{G,E'\} = Cor\{G',E\} = Cor\{G,G'\}$. The number of unknowns would still exceed the number of conditions because there is no a priori information about the magnitude, or even the sign, of the terms I and J.

The genotype-environment covariances are negligible if, and only if, the genotypic and environmental variables are statistically independent to a high degree of approximation. If this is true, then the joint-frequency function (Φ) assumes the form

$$\Phi(x,y) = \Phi_1(x)\ \Phi_2(y) \qquad (9)$$

In these circumstances, not only do all genotype-environment covariances vanish, but so do the covariances $Cov\{G,R\}$, $Cov\{G,R'\}$, $Cov\{E,R\}$, and so forth in Eqs. 6 and 8 (*10*). Equations 5 and 7 then reduce to

$$Var\{P\} = Var\{G\} + Var\{E\} + Var\{R\} \quad (10a)$$

$$Cov\{P,P'\} = Cov\{G,G'\} + Cov\{E,E'\} + Cov\{R,R'\} \quad (10b)$$

This remarkable simplification of the terms *I* and *J* in Eqs. 5 and 7 is a consequence of Fisher's definitions of *G* and *E* (*9*). It does not occur for other definitions. An equally important consequence of these definitions is that they minimize the expectation of R^2 (*11*). That is, they provide the best possible additive representation of the phenotypic value. Both properties hold only so long as the genotypic and environmental variables are independent of one another.

The term $Cov\{R,R'\}$ in Eq. 10b vanishes if either the genotypes or the environments of the two subpopulations are statistically independent. It vanishes, in particular, for monozygotic twins reared in uncorrelated environments and for genetically unrelated individuals reared in identical environments. For these two special cases, respectively, the equations are

$$Cov\{P,P'\} = Var\{G\} \quad (11a)$$

$$Cov\{P,P'\} = Var\{E\} \quad (11b)$$

If all the subpopulations under consideration are fair samples of the parent population, then Eqs. 10a, 11a, and 11b are three equations for three unknowns, and $Var\{G\}$, $Var\{E\}$, and $Var\{R\}$ can all be estimated. If only the phenotypic variance and the variance for separated monozygotic twins are known, their difference provides an upper limit for the environmental variance.

Under certain assumptions, the genotypic covariance between related individuals can be related to the genotypic variance. Consider an equilibrated population of randomly mating diploids, effectively infinite in size so that inbreeding may be neglected, and assume that the trait under consideration is specified by an arbitrary number of unlinked genes, each of which may exist in an arbitrary number of statistically independent allelic states. Under these assumptions, *G* may be written in the form

$$G = \sum_{r,s} G(r,s) \quad (12)$$

where *r* and *s*, in the sum indicated by Σ, take on all integer values (and the value zero) consistent with the condition $r + s \leq n$, the number of pairs of homologous loci specifying the trait. The ordered pair (r,s) specifies a configuration consisting of *r* alleles at nonhomologous loci and *s* pairs of alleles at homologous loci. $G(r,s)$ denotes the sum of all contributions to *G* resulting from allelic configurations of the type (r,s):

$$G(r,s) = \Sigma g(i_1, \ldots, i_r; j_1, \ldots, j_s) \quad (13)$$

where *g* is a random variable depending on the indicated loci and the sum runs over all sets of indices specifying distinct configurations of the type (r,s). In particular, $G(1,0) \equiv G_A$ is the so-called additive contribution to *G*, a sum of $2n$ individual allelic contributions (each represented by a random variable); $G(0,1) \equiv G_D$ is the dominance contribution, a sum of *n* contributions from pairs of homologous loci; $G(2,0)$ is an epistatic contribution, made up of contributions from pairs of nonhomologous loci; and so on. By virtue of the assumptions concerning statistical independence of allelic states, all the random variables that figure in the decompositions (Eqs. 12 and 13) are statistically independent. The variance of *G* is accordingly given by

$$Var\{G\} \equiv \sigma_G^2 = \sum_{r,s} \sigma_{rs}^2$$

$$\sigma_{rs}^2 \equiv Var\{G(r,s)\} = \Sigma\, Var\{g(i_1, \ldots, i_r; j_1, \ldots, j_s)\} \quad (14)$$

where σ_{rs}^2 denotes the variance of the random variable $G(r,s)$. The genotype covariance between individuals related in a known way is given by

$$Cov\{G,G'\} = \sum_{r,s} P(r,s)\sigma_{rs}^2 \quad (15)$$

where $P(r,s)$ is the probability of coincidence between homologous allelic configurations of type (r,s) in the related individuals. In particular, $P(1,0)$ is the probability that a given allele in genome 1 also occurs (at one of the two corresponding loci) in genome 2; $P(0,1)$ is the probability that a given pair of alleles at homologous loci in genome 1 also occurs in genome 2; and so on. From the assumptions about statistical independence, it follows at once that

$$P(r,s) = P_1^r P_2^s \quad (16)$$

where

$$P_1 \equiv P(1,0) = \frac{1}{2}(\phi + \phi')$$

$$P_2 \equiv P(0,1) = \phi\phi' \quad (17)$$

Here ϕ and ϕ' denote the probability of receiving the same gene at a given locus by way of the father and the mother respectively. These results were first obtained by Kempthorne (*12*).

If the two subpopulations are characterized by statistically independent environmental variables, Eq. 10b reduces to

$$Cov\{P,P'\} = \sum_{r,s} P(r,s)\sigma_{rs}^2 \quad (18)$$

For the series in Eq. 14 to be useful in practice, one must be able to truncate it after a number of terms no greater than the number of distinct phenotypic covariances that can be measured. For example, it may be possible to represent the genotypic covariance between parent and offspring or between half-sibs by the leading term in Eq. 15 (in both cases $P_2 = 0$) and thereby estimate $\sigma_{10}^2 = \sigma_A^2$, the additive component of the genotypic variance. The ratio σ_A^2/σ_P^2, called the narrow heritability, determines the rate at which phenotypic changes can be achieved through artificial selection (*13*).

A Conventional Application of Heritability Analysis

Before considering the applicability of the theory just sketched to IQ scores, I will examine some conventional heritability estimates that, as was first pointed out by Lerner (*14*), have an important bearing on the analysis of IQ data. Figure 2 shows heritability estimates, derived from different kinds of correlations, for four commercially important traits in dairy cattle. The primary data show a high degree of internal consistency, in that investigators at different test stations report similar measured correlations. Nevertheless, heritability estimates derived for the same trait from different kinds of data differ substantially and systematically.

Heritability estimates derived from half-sib correlations calculated from field data (indicated by open circles in Fig. 2) are systematically and substantially lower than estimates derived from half-sib correlations calculated from test-station data (indicated by ×'s in Fig. 2). Lerner (*14*) has pointed out that these differences afford a direct and striking illustration of the importance of genotype-environment covariance and genotype-environment interaction.

The fact that environments at test stations are randomized (made random with respect to genotype) greatly reduces, if not entirely eliminates, both effects. (Genotype-environment correlation may be expected to reduce half-sib correlation because it tends to amplify the genotypic effects of the maternal contribution to the genotype of the calf; in dairy cattle, the maternal contributions to the genotypes of half-sibs are uncorrelated.)

For each of the four traits, the heritability estimate derived from separated monozygotic twins is close to unity. This finding can be reconciled with the comparatively low estimates derived from half-sib correlations in two ways:

1) If environmental randomization has, in fact, effectively eliminated correlations between genotypic and environmental variables, so that Eqs. 10a and 11a are applicable, the correlations between separated monozygotic twins are estimates of broad heritability, and the difference $1 - h^2$ provides an upper limit for e^2, the environmental fraction of the phenotypic variance. The half-sib correlations, on the other hand, yield estimates of narrow heritability, and the differences between these estimates and those derived from split monozygotic pairs represent the nonadditive genotypic contribution to h^2.

2) If, however, the environmental value (E) contains a sizable component correlated with the animals' genotype, Eqs. 10a and 11a do not apply. Such a component would be present if the environmental variables were imperfectly controlled, allowing the animals to exercise a modicum of environmental selection—with respect to diet, for example. To the extent that such selection was genetically determined, it would give rise to a "hidden" genotype-environment correlation. And since separated monozygotic twins would tend to select the same environments (to the extent permitted by the controls), their environments would be correlated. In these circumstances, a high phenotypic correlation between monozygotic twins would not entail a correspondingly low value for the environmental component of the phenotypic variance. Hidden genotype-environment interaction would similarly inflate estimates of other phenotypic covariances.

In the absence of additional information, it does not seem possible to choose unequivocally between these explanations, or even to be sure that gene-gene interactions and hidden genotype-environment correlations are jointly responsible for the systematic differences under discussion. Explanation 1 is perhaps the less plausible of the two because it requires the nonadditive genotypic variance to be greatest when the additive genotypic variance is least and because it fails to explain why their sum is nearly the same for traits whose narrow heritabilities span a wide range of values. Until explanation 2 can be ruled out, it must be assumed that estimates of broad heritability based on phenotypic correlations between separated monozygotic twins may be grossly unreliable.

IQ Scores as Measurements of a Phenotypic Character

Before examining the applicability of heritability analysis to IQ scores, one must ask whether such scores can legitimately be assimilated to measurements of a phenotypic character. Tests of IQ differ from measurements of conventional phenotypic characters in two important respects.

In the physical and biological sciences, every measurable quantity is defined in the context of a definite theoretical structure that, in general, serves to generate a variety of distinct operational definitions. For example, the physical geometry of rigid bodies provides the basis for several operationally distinct ways of measuring length and distance. Electromagnetic theory provides a second set of operational definitions of distance and length; the theory of sound a third set; and so on. The flexibility conferred by the existence of operationally distinct ways of measuring the same quantity is important because it makes possible the detection and eventual elimination of systematic errors; and the existence of distinct operational definitions implies an underlying theoretical structure.

The definition of IQ has no theoretical context or substratum. Tests of IQ measure what they measure. They are precisely analogous to physical readings made with a black box—a device whose internal working is unknown. Because we do not know what an IQ test or a black box is supposed to measure or how it works, we cannot know to what extent measurements carried out on different subjects are comparable or to what extent they are influenced by extraneous factors. Thus IQ scores contain uncontrollable, systematic errors of unknown magnitude.

This helps to explain why different investigators frequently report such widely differing estimates of the same IQ correlation. For example, reported estimates of the parent-child correlation range from .2 to .8, while estimates of the correlation between same-sex dizygotic twins range from .4 to .9 (*15*). According to Jensen (*16*), there are no objective criteria (other than sample size) for weighting discrepant estimates of the same correlation.

Because the definition of IQ is purely instrumental, it fails to confer the most essential attribute of a scientific measurement—objectivity. To measure a subject's Stanford-Binet IQ, one must administer a specific test in a specific way under specific conditions. By contrast, a well-equipped physics laboratory does not need to have replicas of the standard meter and the standard kilogram to measure length and mass, and the physicist or biologist is free to devise his own techniques for measuring such quantities. Systematic discrepancies between measurements of the same quantity are never ignored in the physical and biological sciences, because they signal the presence of unsuspected systematic errors or of defects in the theory underlying the measurements.

IQ scores also differ from conventional measurements in that they have no strict quanitative meaning. The IQ is an index of rank order on a standardized test, expressed according to a convenient but essentially arbitrary convention (*17*). In effect, the intervals of the IQ scale are chosen in such a way as to make the frequency distribution of test scores in a reference population approximately normal, but other methods of defining the scale could claim equal prior justification.

These considerations show that IQ scores are not phenotypic measurements in the usual sense. This is not to say that they have no scientific value or practical utility, or that certain aspects of intelligent behavior cannot, in principle, be adequately defined, quantified, and measured. It is even conceivable that valid measurements of some well-defined behavioral trait would rank subjects in roughly the same way as IQ tests and that they would turn out to be normally distributed. In these circumstances, IQ scores would indeed approximate valid measurements of a phenotypic trait,

although they would continue to be afflicted by uncontrollable systematic errors of unknown magnitude. For the moment, however, I ignore these shortcomings of the primary data, and turn to the methods available for analyzing them.

Heritability of IQ

In applications of heritability analysis to metric (continuously variable) characters in plants and animals, at least some of the relevant genetic and environmental factors are under experimental control. In particular, plant and animal geneticists can minimize genotype-environment correlation by randomizing environments. As I have shown, this step is indispensable for the application of heritability analysis. Unless Eqs. 5 and 7 can be replaced by Eqs. 10a and 10b, there is no hope of disentangling the genotypic and environmental contributions to phenotypic variances. Recent discussions of the applicability of heritability analysis to IQ scores have failed to stress this point. The applicability of heritability analysis does not, as is commonly assumed, hinge on the smallness of the interaction term (R) relative to the terms G and E in Fisher's decomposition of the phenotypic value. In fact, one may reasonably assume on biological grounds that genotype-environment interaction makes a substantial contribution to the phenotypic value of every phenotypically plastic trait, except in populations where the ranges of genetic and environmental variation are severely restricted. Even so, heritability analysis can be applied to phenotypically plastic traits, provided that the relevant genetic and environmental variables are statistically uncorrelated. When this condition is not satisfied, the contributions of interaction to phenotypic variances and covariances cannot, in general, be separated from the contributions of genotype and environment, and heritability analysis cannot, therefore, be applied meaningfully.

In adult subpopulations, IQ and environment are well known to be more or less strongly correlated. Since differences in IQ are undeniably related to genetic differences (although not, perhaps, in a very simple way), one may safely assume that genotype-environment correlation is significant in adult subpopulations and in subpopulations composed of children reared by their biological parents or by close relatives. Hence, no valid estimate of IQ heritability can be based on data that refer to such subpopulations.

Yet data of precisely this kind make up the bulk of the available material, and many published heritability estimates have been based on them. Burt (*18*), Jensen (*1*), and Herrnstein (*4*), for example, all cite kinship correlation data as evidence for a high value of h^2.

These authors, among others, rely especially on IQ correlations between separated monozygotic twins. Such correlations, however, are highly sensitive to distortion by genotype-environment correlation. As I have shown, significant distortion may occur even under experimental conditions in which the controllable aspects of the environment have been randomized, because selection of "microenvironments" is genetically influenced. The development of human cognitive skills is presumably even more sensitive to such selection than the development of physiological characters of dairy cattle. In published studies of separated monozygotic twins, no serious attempts have been made to minimize the effects of genotype-environment interaction—something that would be very difficult to do under the best of circumstances. For example, in the largest and the most homogeneous of the four major twin studies, that of Burt (*18*), one member of each of the 53 pairs included in the study was reared by his or her natural parent (*19*).

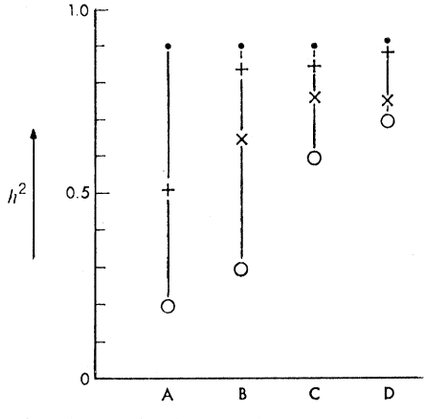

Fig. 2. Heritability estimates for four metric traits in dairy cattle: (A) calf weight and measure, (B) milk yield, (C) fat percentages, and (D) body weight and measure at 2 years. The estimates are derived from comparisons among separated monozygotic twins (closed circle), monozygotic and dizygotic twins (cross), half sisters at testing stations in which environment is randomized (×), and half sisters at field stations (open circle). [Source: Donald (*25*), reproduced by Lerner (*14*, p. 412)]

Since reliable inferences about the heritability of behavioral traits cannot be drawn from data referring to subpopulations of adults or of children reared by their biological parents or by close relatives, one must turn to studies in which the subjects are adopted children. For example, phenotypic correlations between half-sibs reared in foster homes would yield estimates of narrow heritability analogous to those indicated by ×'s in Fig. 2—provided that there was no selective placement and that the range and distribution of relevant environmental factors was the same for the subpopulation of foster homes as for the reference population. Analogous data on siblings would be somewhat less useful, because gene-gene interactions—in particular, the dominance covariance—could contribute substantially to the phenotypic covariance. However, no suitable data for either siblings or half-sibs seem to exist (*6*).

Finally, the IQ correlation between unrelated, randomly selected children reared in the same foster home could, in principle, provide an estimate of e^2 (Eq. 11b). The phenotypic covariance might substantially underestimate e^2, however, because the environments of children reared together are never identical. Age differences may give rise to significant "macroenvironmental" differences, while differences in genotype may give rise to significant "microenvironmental" differences, as discussed above.

Jencks *et al.* (*6*) cite four studies of pairs of unrelated foster children reared together. The reported IQ (Stanford-Binet) correlations, corrected for unreliability and restriction of range, are as follows: .17 (10 cases), .29 (21 cases), .46 (93 cases), and .72 (41 cases). Using Fisher's method (*20*) to combine these estimates, one arrives at the value $r = .5 \pm .05$. The differences between the individual estimates of r, although large, are not statistically significant, according to the test described by Fisher (*20*). For the reasons explained above, the environments of unrelated foster children reared together may be significantly different. On the other hand, selective placement may introduce a correlation between the genotypes of the foster children. The two effects work in opposite directions, and neither can be reliably estimated. Making the not-very-safe assumption that they cancel, one arrives at the highly tentative

estimate, $e^2 \simeq .5 \pm .05$. This estimate refers to the purely environmental fraction of the phenotypic variance. Of greater relevance to the social and educational issues mentioned at the beginning of this article is the nongenetic fraction of the phenotypic variance, given by $1 - h^2 = e^2 + i^2$, where $i^2 = Var\{R\}/Var\{P\}$ is the fraction of the phenotypic variance relating to interaction. (As noted earlier, this decomposition of the phenotypic variance presupposes that genotype-environment correlations are absent.) Since there are no available data that would permit an independent estimate of the interaction contribution, the above estimate of e^2 provides only a lower limit for $1 - h^2$.

Thus the available data for unrelated foster children raised together yield the following, highly tentative estimate of the broad heritability:

$$0 \leqq h^2 \lessapprox .5 \qquad (19)$$

It is important to notice that this estimate refers specifically to the subpopulation of unrelated foster children reared together. [For this subpopulation, it is consistent with the estimates of Jencks *et al.* (*6*).] It does not apply to populations composed of children reared by their biological parents or by near relatives, because, as shown above, both the conceptual and operational definitions of heritability break down in the presence of significant genotype-environment correlation. This explains why Jencks *et al.* (*6*) were unable to reconcile their heritability estimate based on unrelated foster children reared together with estimates derived from other kinds of IQ data. The present considerations show that the estimates based on other kinds of data are, in a strict sense, meaningless.

Systematic effects of genotype-environment correlation are by no means the only obstacles to meaningful analyses of IQ correlations. Additional serious difficulties arise from:

1) *Ignorance of the specific environmental factors affecting cognitive development.* Because social scientists have not yet identified the specific environmental factors most relevant to cognitive development, they are unable to assess environmental similarities or differences objectively, even at a qualitative level. For example, some authors (*1*, p. 52; *4*, p. 55) have assumed that the within-pair environmental differences between separated monozygotic twins in Burt's study (*18*) are representative of those between randomly selected subjects of the same age. This assumption is based on a reported lack of correlation between the occupational statuses of the biological and adoptive fathers. As far as I know, no evidence has been adduced to support the implied assumption that the occupational status of the father plays a crucial role in cognitive development. Current studies suggest that specific kinds of mother-child interaction during infancy and early childhood do play a significant role in cognitive development (*21*), but the study of such interactions is still in a primitive stage.

2) *Nonrandom mating.* If one wishes to analyze phenotypic correlations between related persons other than monozygotic twins, one must allow for assortative mating. This introduces two more unknowns into the analysis: the environmental and the phenotypic covariances between mates. In the absence of reliable assessments of relevant environmental factors, the environmental covariance is not measurable, so the only additional datum would be the phenotypic covariance. Thus assortative mating introduces a further unknown into an already top-heavy analysis. Moreover, Eq. 15 (for the genotypic covariance) applies only to populations with random mating, and an appropriate generalization of it that allows for assortative mating has not yet, to my knowledge, been made.

3) *Gene-gene interactions.* The possibility of evaluating the narrow heritability of a trait hinges on how rapidly the series in Eq. 15 converges, the number of unknown components of the genotypic variance one can hope to estimate being limited by the number of measured kinship correlations. Now, for a trait specified by n pairs of genes, σ^2_{rs} is made up of $2^r n!/r!s!\,(n\text{-}r\text{-}s)!$ separate contributions. This suggests that, as n increases, the relative importance of gene-gene interactions of a given kind also increases: the greater the number of genes contributing to the specification of a trait, the more likely it is that nonadditive genetic effects will play an important role. These considerations raise the possibility that human intelligence, however it may be defined, could depend on the total genotype in a manner too complex to permit the application of conventional heritability analysis. In any case, one may reasonably expect to find substantial differences between the broad and narrow heritabilities of cognitive traits (*22*).

In view of the difficulties discussed above, studies of unrelated foster children reared together and of half-sibs reared in foster homes seem to offer the only realistic prospects for estimating, respectively, the environmental fraction of the IQ variance and the narrow heritability.

Alternative Quantitative Approaches

Heritability analysis was devised to help answer one of the central practical questions of plant and animal genetics: Under given environmental conditions, how rapidly can systematic changes in a metric character be produced by artificial selection? Fisher's "fundamental theorem of natural selection" (*23*) implies that the rate of evolution in question is proportional to the additive genotypic variance and hence to the narrow heritability of the character. For obvious reasons, this question—and, therefore, the value of the narrow heritability—is not of comparable importance for human behavioral traits. On the other hand, the phenotypic plasticity of human behavioral traits is of considerable interest both to geneticists and to students of human behavior. Estimates of the broad heritability—or, better still, of e^2—do tell us something about the sensitivity of a trait to environmental variation, but they throw little light on what are perhaps the most important questions: To what sorts of environmental changes are behavioral traits most sensitive? To what extent can cognitive performance be improved by appropriate forms of environmental intervention? How do genetic differences affect levels of cognitive performance attainable under optimal environmental conditions? These and similar questions are not of less scientific interest than those to which heritability estimates provide answers; and they are considerably more relevant to the educational, social, and political issues mentioned at the beginning of this article.

It may turn out to be less difficult to find semiquantitative or even quantitative answers to such questions than to obtain reliable heritability estimates. Notable progress along these lines has already been made. The remarkable achievements of the Milwaukee Project (*24*), to cite a single example, afford a direct and dramatic demonstration of the efficacy of appropriate environmental modifications in accelerating

cognitive development. In this study, now in its sixth year, a comprehensive family intervention program produced a sustained, 30-point difference in IQ between an experimental group and a control group, each composed of 20 randomly selected children of mothers with tested IQ's of under 75. Over a 5-year period, the average IQ of the experimental group remained close to 125. Children in the experimental group were evaluated through independent tests administered by psychologists not connected with the study.

The methodological difficulties of intervention studies should not be minimized. Nevertheless, the nature of the questions such studies are trying to answer makes these difficulties inherently less formidable than those besetting the application of conventional heritability analysis to IQ scores. For example, the shortcomings of IQ as a phenotypic measurement, although they cast serious doubt on the meaningfulness of heritability estimates, do not impair the usefulness of IQ tests for assessing differences in cognitive performance between an experimental group and a control group in studies like that of Heber and his colleagues.

IQ and Race

Jensen (*2*) and others have argued that reported differences in average IQ between black and white children are probably attributable in part to systematic genetic differences. Jensen explicitly states the "rule of inference" used to draw this conclusion (*16*, p. 438): "The probability that a phenotypic mean difference between two groups is in the same direction as a genotypic mean difference is greater than the probability that the phenotypic and genotypic mean differences are in opposite directions." This rule fails, however, when systematic effects whose magnitude cannot be estimated are known to contribute to the phenotypic mean differences. (Unfortunately, these are the only circumstances in which the rule might be useful.) In order to estimate the probabilities mentioned in the rule, one would need to estimate the probability that the observed phenotypic mean difference exceeds the contribution of systematic effects—which, by assumption, is impossible.

Among the relevant systematic differences between blacks and whites are cultural differences and differences in psychological environment. Both influence the development of cognitive skills in complex ways, and no one has succeeded in either estimating or eliminating their effects. "Culture-free" tests deal with this problem only on the most superficial level, for culture-free and "culture-bound" aspects of cognitive development are inseparable. The difficulties cannot be overcome by refined statistical analyses. As long as systematic differences remain and their effects cannot be reliably estimated, no valid inference can be drawn concerning genetic differences among races.

Precisely the same arguments and conclusions apply to the interpretation of IQ differences between socioeconomic groups.

Summary and Conclusions

Estimates of IQ heritability are subject to a variety of systematic errors.

The IQ scores themselves contain uncontrollable, systematic errors of unknown magnitude. These arise because IQ scores, unlike conventional physical and biological measurements, have a purely instrumental definition. The effects of these errors are apparent in the very large discrepancies among IQ correlations measured by different investigators.

Genotype-environment correlations, whose effects can sometimes be minimized, if not wholly eliminated, in experiments with plants and animals, are nearly always important in human populations. The absence of significant effects arising from genotype-environment correlations is a necessary condition for the applicability of conventional heritability analysis to phenotypically plastic traits. When this condition fails, no quantitative inferences about heritability can be drawn from measured phenotypic variances and covariances, except under special conditions that are unlikely to be satisfied by phenotypically plastic traits in human populations.

Inadequate understanding of the precise environmental factors relevant to the development of specific behavioral traits is an important source of systematic errors, as is the inability to allow adequately for the effects of assortative mating and gene-gene interaction.

Systematic cultural differences and differences in psychological environment among races and among socioeconomic groups vitiate any attempt to draw from IQ data meaningful inferences about genetic differences.

Estimates based on phenotypic correlations between separated monozygotic twins—usually considered to be the most reliable kind of estimates—are vitiated by systematic errors inherent in IQ tests, by the presence of genotype-environment correlation, and by the lack of detailed understanding of environmental factors relevant to the development of behavioral traits. Other kinds of estimates are beset, in addition, by systematic errors arising from incomplete allowance for the effects of assortative mating and from gene-gene interactions. The only potentially useful data are phenotypic correlations between unrelated foster children reared together, which could, in principle, yield lower limits for e^2. Available data indicate that, for unrelated foster children reared together, the broad heritability (h^2) may lie between 0.0 and 0.5. This estimate does *not* apply to populations composed of children reared by their biological parents or by near relatives. For such populations the heritability of IQ remains undefined.

The only data that might yield meaningful estimates of narrow heritability are phenotypic correlations between half-sibs reared in statistically independent environments. No useful data of this kind are available.

Intervention studies like Heber's Milwaukee Project afford an alternative and comparatively direct way of studying the plasticity of cognitive and other behavioral traits in human populations. Results obtained so far strongly suggest that the development of cognitive skills is highly sensitive to variations in environmental factors.

These conclusions have three obvious implications for the broader issues mentioned at the beginning of this article.

1) Published analyses of IQ data provide no support whatever for Jensen's thesis that inequalities in cognitive performance are due largely to genetic differences. As Lewontin (*8*) has clearly shown, the value of the broad heritability of IQ is in any case only marginally relevant to this question. I have argued that conventional estimates of the broad heritability of IQ are invalid and that the only data on which potentially valid estimates might be based are consistent with a broad heritability of less than 0.5. On the other hand, intervention studies, if their findings prove to be replicable,

would directly establish that, under suitable conditions, the offspring of parents whose cognitive skills are so poorly developed as to exclude them from all but the most menial occupations can achieve what are regarded as distinctly high levels of cognitive performance. Thus, despite the fact that children differ substantially in cognitive aptitudes and appetites, and despite the very high probability that these differences have a substantial genetic component, available scientific evidence strongly suggests that environmental factors are responsible for the failure of children not suffering from specific neurological disorders to achieve adequate levels of cognitive performance.

2) Under prevailing social conditions, no valid inferences can be drawn from IQ data concerning systematic genetic differences among races or socioeconomic groups. Research along present lines directed toward this end—whatever its ethical status—is scientifically worthless.

3) Since there are no suitable data for estimating the narrow heritability of IQ, it seems pointless to speculate about the prospects for a hereditary meritocracy based on IQ.

References and Notes

1. A. R. Jensen, *Harv. Educ. Rev.* **39**, 1 (1969).
2. ———, *Genetics and Education* (Harper & Row, New York, 1972).
3. ———, *Educability and Group Differences* (Methuen, London, 1972).
4. R. J. Herrnstein, *Atl. Month.* **228**, 43 (1971).
5. ———, *IQ and the Meritocracy* (Little, Brown, Boston, 1973).
6. C. Jencks *et al.*, *Inequality: A Reassessment of the Effects of Family and Schooling in America* (Basic Books, New York, 1972), appendix A, pp. 266–319.
7. A similar argument was given as long ago as 1933 by Hogben [cited by C. H. Waddington, *The Strategy of the Genes* (Allen & Unwin, London, 1957), pp. 92–94].
8. R. C. Lewontin, *Bull. At. Sci.* **26**, 2 (March 1970); in *The Fallacy of IQ*, C. Senna, Ed. (Third press, New York, 1973), pp. 1–17.
9. R. A. Fisher, *Trans. R. Soc. Edinb.* **52**, 399 (1918). In this fundamental paper, Fisher expresses the genotypic value (G) of a trait as the sum of additive allelic contributions ($G_1, G_2, \ldots G_n$) and dominance contributions ($G_{12}, G_{13}, \ldots$) in such a way as to minimize the contribution of dominance to the genotypic variance (pp. 404, 415). The decomposition (Eq. 4) of the phenotypic value (P) into additive genotypic and environmental values (G and E) and a residual interaction term (R) is formally analogous: E (in the formula $E = E_1 + E_2 + E_{12}$) corresponds to P (in the formula $P = G + E + R$), E_1 to G, E_2 to E, and E_{12} to R. The formal analog of the requirement of statistical independence between genotype and environment (Eq. 9) is Fisher's assumption that genes assort independently. Thus the effects of linkage are formally analogous to those of genotype-environment correlation.
10. Consider first the case of a single genotypic variable $x \equiv x_1$, and a single environmental variable $y \equiv y_1$. That the covariances in question do indeed vanish follows immediately from the power-series representations of G, E, and R:

$$G = \sum_k \frac{1}{k!} \left(x^k - \mathcal{E}\,(x^k)\right) P_{k,0}(0,0)$$

$$E = \sum_j \frac{1}{j!} \left(y^j - \mathcal{E}\,(y^j)\right) P_{0,j}(0,0)$$

$$R = \sum_{r>1} \sum_{s>1} \frac{1}{(r+s)!} \left[x^r y^s - \mathcal{E}(x^r)\ \mathcal{E}(y^s)\right] P_{r,s}(0,0)$$

where

$$P_{r,s}(0,0) = \left[\frac{\partial^{r+s}}{\partial x^r \partial y^s} P(x,y)\right]_{x=y=0}$$

These formulas follow from the Taylor expansion of P (assumed to be valid for all x,y) and the defining conditions (*2*). In the general case, the proof is notationally more complex but essentially the same.
11. For notational simplicity, let $x \equiv x_1$, $y \equiv y_1$. The most general form of R (assuming that it has derivatives of all orders at $x = y = 0$) is

$$R = R^*(x,y) + f(x) + g(y)$$

where R^* is the expression for R given earlier (*10*), g is a random variable depending on the indicated loci, and $\mathcal{E}\{f\} = \mathcal{E}\{g\} = 0$. By expanding f and g, one finds that $\mathcal{E}\{R^* f\} = \mathcal{E}\{R^* g\} = 0$. Hence

$$\mathcal{E}\{R^2\} = \mathcal{E}\{R^{*2}\} + \mathcal{E}\{f^2\} + \mathcal{E}\{g^2\} \geq \mathcal{E}\{R^{*2}\}$$

→ O. Kempthorne, *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **143**, 103 (1954).
13. Narrow heritability is also the kind of heritability that is relevant to Herrnstein's (*4*) meritocracy argument.
14. I. M. Lerner, *Evol. Biol.* **6**, 399 (1972).
15. L. Erlenmeyer-Kimling and L. F. Jarvik, *Science* **142**, 1477 (1963).
16. A. R. Jensen, *Cognition* **1**, 427 (1973); see also D. Layzer, *ibid.*, p. 453.
17. This point has been made by S. S. Stevens, *Science* **103**, 677 (1946).
18. C. Burt, *Br. J. Psychol.* **57**, 137 (1966).
19. Jencks *et al.* (*6*) recognized that estimates of h^2 derived from twin studies must make allowance for genotype-environment correlation, and they have, in fact, attempted to do this. However, once the effects of genotype-environment correlation are admitted to be significant, the problem of estimating h^2 becomes mathematically indeterminate: the number of unknowns exceeds the number of available conditions. The analysis of Jencks *et al.*, although considerably more elaborate and realistic than previous analyses of the same data, depends on plausible but ad hoc assumptions to close the gap between the number of unknowns and the number of conditions.
20. R. A. Fisher, *Statistical Methods for Research Workers* (Oliver & Boyd, Edinburgh, ed. 13, 1958), p. 204.
21. U. Bronfenbrenner, *Is Early Intervention Effective?* (Cornell Univ. Press, Ithaca, N.Y., in press).
22. Thus Jensen's estimate of broad heritability, whatever its validity, is not directly relevant to the argument by which Herrnstein (*4*) predicts the emergence of an hereditary meritocracy. This is not to say that the argument is free from other serious defects.
23. R. A. Fisher, *The Genetical Theory of Natural Selection* (Dover, New York, ed. 2, 1958), p. 22.
24. R. Heber, H. Garber, S. Harrington, C. Hoffman, "Rehabilitation of families at risk for mental retardation" (unpublished research report). Some writers [for example, Jensen (*16*, p. 427)] have expressed skepticism concerning the reported gain of IQ because of its size. Recent findings by R. B. McCall, M. I. Appelbaum, and P. S. Hogarty [*Monographs of the Society for Research in Child Development* **38**, 1 (1973)] may help to put those of Heber and his colleagues in perspective: "Normal home-reared middle-class children change in IQ performance during childhood, some a substantial amount . . . the average individual's range of IQ between 2½ and 17 years of age was 28.5 IQ points, one of every three children displayed a progressive change of more than 30 points, and one in seven shifted more than 40 points. Rare individuals may alter their performance as much as 74 points" (p. 70).
25. H. P. Donald, in *Proceedings of the Tenth International Congress of Genetics* (Univ. of Toronto Press, Toronto, 1959), vol. 1, pp. 225–235; reproduced by Lerner (*14*, p. 412).

26. I am indebted to L. Cavalli-Sforza, E. R. Dempster, and I. M. Lerner for helpful comments, criticism, and suggestions.