Psychometrics for Educational Debates
Edited by L. J. Th. van der Kamp, W. F. Langerak and D. N. M. de Gruijter

87

Genetic–Environment Covariation in Human Behaviour Genetics†

Joseph M. F. Jaspars and Jan A. de Leeuw

A CENTURY OF BIOMETRICAL GENETICS

Exactly a century before Jensen (1969) published his controversial paper on genetics, intelligence, and scholastic achievement in the *Harvard Educational Review*, the first quantitative inquiry into the laws and consequences of hereditary ability by Galton (1892) appeared. It is sometimes suggested that behaviour genetics came of age only recently. Jensen's paper certainly does not provide many clues to the long and controversial history of the genetic determination of human abilities. We want to point out, however, that the developments which have taken place in behaviour genetics are mainly of a technical nature, i.e. the refinement of statistical procedures for the analysis of observed variation in behaviour (see Eysenck, this book). Conceptually there is a great deal of similarity between the earliest studies and recent reanalyses of material which was collected a long time ago. The model on which these analyses are based has not changed drastically since Fisher's classical paper in 1918 and one might even argue, since Mendel and Galton.

It is probably not clear at all to many readers of Jensen's review article, that his discussion of the available evidence is firmly based in this long biometrical tradition which is characterized by a very specific theoretical orientation. We believe that this orientation is shown most clearly in the conception of the environment in biometrical genetics. By this we do not merely want to indicate the minor role which has often been assigned to the relevance of environmental variation for the explanation of variation in behaviour, but also the conception of the nature of the environment itself. We will argue that the

† The ideas discussed in this paper owe a great deal to the participants of the graduate seminar on 'Behaviour Genetics and Social Equality' at Leyden University in 1976–77.

environment has been conceived of at first essentially as a random variable and later as a passive and static factor. These notions have important consequences for the results one obtains in genetic analyses. We will present, therefore, some models on which the environment is also seen as active and reactive and elaborate the possible consequences of this approach for the interpretation of differences in intelligence and educational performance.

GALTON'S LAW OF ANCESTRAL HEREDITY

In *Hereditary Genius* Galton presents two classifications of men; one according to their reputation and another according to their natural gifts. According to Galton, reputation is a good indication of a person's natural ability, because the environment can never repress the rise to eminence for men of extraordinary talents. Curiously enough the data presented by Galton in *Hereditary Genius* do not support this assertion of complete environmental insignificance.

In Table 1 we have reproduced the summary of the results obtained by Galton and presented in *Hereditary Genius*. There are several interesting features to be noticed in Table 1. The first is that the percentage of eminent sons having eminent fathers is not the same as eminent fathers having eminent sons. A simple genetic model would make the same predictions in both cases. The discrepancy in Galton's case suggests that either assortative mating and/or environmental factors play a role in the covariation of eminence

Table 1 Percentages of eminent kinsmen in various degrees of family relationships (after Galton, 1892, p. 308)

	Co	S	T	C	D
Father	47	26	31	100	31
Brother	50	47	41	150	27
Son	31	60	48	100	48

Co = commanders, S = scientists, T = total, C = number of individuals in each degree to 100 men, D = 100T/C.

The percentages in the cells indicate the conditional probability of having an eminent father, brother, or son, given that the person who is considered is eminent himself. Columns 1 and 2 give Galton's data for military commanders and scientists. Column 3 gives the average for all groups mentioned by Galton. Column 4 introduces the connection factor for various kinship relationships. Galton estimates that on the average his subjects had only 1 father, 1.5 brothers, and 1 son.

between fathers and sons, or that Galton has underestimated the number of eminent fathers.

It appears that Galton was well aware of the fact that these data did not confirm his expectations. He notices first that in the case of military commanders the conditional probabilities for the father–son relationship are reversed (see Table 1). Galton explains this irregularity by pointing out that military commanders begin their active careers in youth and therefore, if married at all, are mostly away from their wives. Galton does not offer the opposite explanation for the fact that in the group of scientists and artists the percentage of eminent sons is much higher than the average. In both cases Galton argues that training and the blood of their mothers can account for this anomaly. That a man should have an able mother for achieving success in science is regarded by Galton as very important, because such mothers will show their children by practice and teaching to follow science as a profession and not to waste their powers on profitless speculations. Apparently, Galton is in this case of the opinion that intelligent mothers develop the inherited abilities of their eminent children by pointing out the right way to develop their talents. This process of parent–infant interaction does imply genetic-environment covariation, but this notion is not taken up by Galton in his law of ancestral heredity.

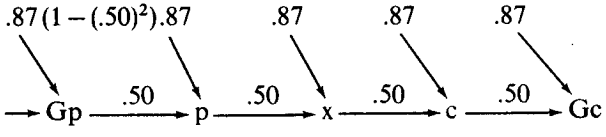
The first formulation of the law of ancestral heredity is in two early papers by Galton (1865) on hereditary talent and character, which are based in part on the same observations as *Hereditary Genius*. In the second of these papers Galton states that 'the share a man retains in the constitution of his remote descendants is inconceivably small. The father transmits on the average one half of his nature, the grand-father one-fourth, the great-grandfather one eighth, the share decreasing step-by-step in a geometrical ratio with great rapidity.' It is not clear how Galton had already reached this conclusion in 1865. However, the argumentation Galton presents in 1889 is mainly statistical in nature and based upon filial regression data for stature, Galton first shows that for stature the ratio of filial deviation to mid-parental deviation from the population mean is $\frac{2}{3}$. Next he shows that there is also, to his own surprise, a regression from the son to the mid-parent of $\frac{1}{3}$. By repeated application of this converse regression Galton, quite incorrectly, arrives at the conclusion that the total bequeathable ancestral property is $1+\frac{1}{3}+\frac{1}{9}+\text{etc.}=\frac{3}{2}$. To find the influence pure and simple of the mid-parent, Galton computes the ratio of the effective and the accumulative heritage $\frac{2}{3}/\frac{3}{2}$ which is approximately $\frac{1}{2}$. Apart from the fact that Galton's reasoning is curious, to say the least, it is quite obvious that Galton's law of ancestral heredity does not specify in any way a biological process or mechanism by which parental and ancestral heritage is transmitted. It describes equally well the way in which parents bequeath material property to their children. Frogatt and Nevin (1971) have pointed out that Galton's law was influenced by the idea that biological inheritance was

analogous to legal inheritance. After a thorough historical study the same authors conclude that it is not certain whether Galton presented his laws as an empirical prediction model for phenotypical resemblance or as an expression of his physiological views of inheritance, which were in part determined by Darwin's theory of pangenesis.

Karl Pearson (1898), who reformulated Galton's law as a linear regression equation, is more explicit in this respect, when he writes that the numerical laws for the intensity of inheritance must first be discovered from wide observation, before plasmic mechanics can be anything but the purest hypothetical speculation. Pearson's version of Galton's law shows perhaps more clearly than anything else the purely phenotypical nature of the law of ancestral heredity. It states that

$$x_0 = \frac{1}{2}\frac{\sigma_0}{\sigma_1}x_1 + \frac{1}{4}\frac{\sigma_0}{\sigma_2}x_2 + \frac{1}{8}\frac{\sigma_0}{\sigma_3}x_3 + \text{etc.}$$

where x_0 is the predicted deviation of an offspring from the generation mean, x_1, x_2 etc. the deviation of the (mid)parent, (mid)grandparent, etc. from the generation mean, and σ_0, σ_1, σ_2 the respective standard deviations. The interesting point of this equation is that it describes a simple recursive system of direct causal influences between successive phenotypical generations where $r_{x_i x_j} = .50$. The path diagram representing the system is simply



where Gp, p, x, c and Gc indicate grandparent, parent etc. The law simply describes the fact that the phenotype of everyone's mid-great-great-grandparent is practically the same, which is indeed the case for IQ as Eaves (1973) has recently shown.

What makes Galton's law a genetical law is a reasoning by analogy. The genetical interpretation of statistical data of family relations in psychological characteristics rests upon a comparison with plant and animal studies or biological data in humans. In one of the earliest publications by K. Pearson (1904) on the inheritance of mental and moral characteristics in man, he shows that the correlations for various physical and mental characteristics of siblings reared together is close to $r = .50$. Pearson attaches almost a mystical importance to the fact that psychical characters in man are inherited to the same extent as the protopodite of the water-flea. What reason is there, Pearson asks rhetorically, for demanding a special evolution of man's mental and moral side. We look upon the universe and wonder. The only alternative explanation

Pearson offers is the most marvellous possibility that with varying degrees of inheritance, some mysterious force always modifies the extent of home influence, until the resemblance of brothers and sisters is brought up to the same intensity. Pearson finds his highly unlikely and concludes that we are forced, I think literally forced, to the conclusion that the physical and psychical characters in man are inherited in their broad lines in the same manner and with the same intensity.

It will be clear that the conclusion of Pearson would lose all its force if the correlations for mental characteristics appeared not to be close to $\bar{r}=.50$. Elderton (1923), who has reviewed the evidence from early studies on the inheritance of intelligence, finds approximately the same average value is Pearson in his first study ($\bar{r}=.485$), but the variation in at least one study is considerable with values ranging from .27 to .67. Even lower correlations were found by Heymans (1927) in the Netherlands. The highest correlation found by Heymans is $\phi=.32$ for mathematical talents, but many correlations are very low. Heymans nevertheless concludes that in general hereditary influence exceeds by far the influence of education because his 'hereditary' coefficients are approximately the same for traits which can and cannot be influenced by environmental factors.

In summary it appears that the earliest work in biometrical genetics by Galton, Pearson, and others can only be considered correlational, phenotypical, and descriptive. No testable biological theory or model was offered for the correlations observed for various family relationships, while environmental explanations were ruled out by comparison of mental and physical characteristics. It is interesting to note that such comparisons are still made in the work of Burt and Jensen, and considered as evidence for the high heritability of population differences in intelligence.

With the rediscovery of Mendel's work a new period in the genetic study of psychological characteristics of humans emerged, because Mendel's theory offered a model from which correlations between relatives could be derived. Although Galton and Pearson had tried to integrate biometrical genetics with Mendel's model, the landmark in this respect is of course Fisher's paper of 1918 in which for the first time a comprehensive correlational approach, based upon Mendel's ideas was presented. It is not our intention to discuss Fisher's paper in detail. We merely want to make clear how environmental influences are taken into account in Fisher's analysis.

Fisher's integration

Fisher's (1918) attempt to interpret the well-established results of biometry in accordance with the Mendelian scheme of inheritance is a test of a purely genetic model in which the environment is introduced only as an uncorrelated random factor. Fisher proceeds from the simple to the complex and assumes at

first random mating and no dominance. The genetic correlations given in many later publications by Burt, Jensen, and Eysenck are based on this model. Fisher's treatment of dominance, epistacy, assortative mating, coupling, and multiple allelomorphism in the middle part of his paper does not change the initial model in the sense that it now includes an environmental component. Fisher discusses the problem of environmental influences only in relation to dominance. As Fisher indicates, he is using the term 'environment' 'formally for arbitrary causes independent of heredity' (1918, p. 420). Apart from this Fisher has shown that under random mating conditions, the effect of dominance will be to reduce the parental correlation twice as much as the fraternal correlation and suggests that a comparison of the two correlations allow us to distinguish between random effects of the environment and dominance. In assessing the relative importance of dominance and environment for data of stature Fisher, however, simply assumes that the environment has no effect, although he observes in the text that random environmental effects might account for something up to 5 per cent of the total variance.

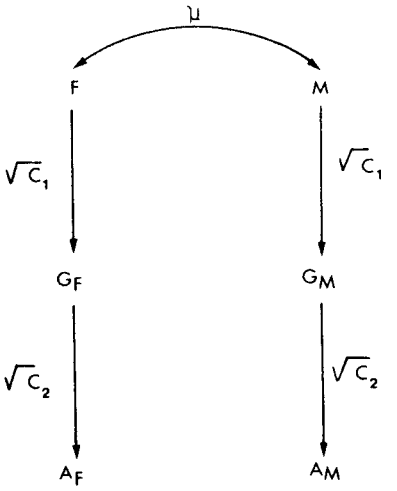
Burt and Howard (1965, p. 115), who have noticed this problem in Fisher's analysis offer a more satisfactory approach, which is basically the one still in use in biometrical genetics. Given the phenotypical correlations for parents and offspring the proportion of variance attributable to additive genetic factors (essential genotypes), assortative mating, dominance, and environment may be estimated using one of Fisher's more complex models. In one of these models:

$$\mu=\frac{A}{C_1C_2}, \qquad p=C_1C_2\left(\frac{1+\mu}{2}\right), \qquad f=\frac{C_1}{4}(1+C_2(1+2A))$$

where μ is the marital correlation of phenotype, A the marital correlation of essential genotypes, p the parent–offspring or parental correlation, and f the fraternal correlation, and:

$$C_1=\frac{V_A+V_D}{V_p} \quad \text{and} \quad C_2=\frac{V_A}{V_A+V_D}$$

where V_A is the additive genetic variance, V_D the variance due to dominance effects, and V_p the phenotypical variance, so that C_1 indicates the effect of dominance and C_2 the effect of the random environment. In this particular model Fisher assumes that the effect of assortative mating originates at the phenotypical level and not, as is the case with inbreeding, at the additive genetical level. Fisher's derivation of the various equations is difficult to follow. Path diagrams can, however, readily be drawn from which the three equations can be derived. Figure 2 presents Fisher's complete model in the form of a path diagram. Path diagrams for marital, parental, and fraternal correlations



$$A = r_{A_F A_M} = \sqrt{C_2}\,\sqrt{C_1}\,\mu\,\sqrt{C_1}\,\sqrt{C_2} = C_1 C_2 \mu$$

where A = correlation between additive genotypes of parents

μ = correlation between phenotypes of parents.

F = phenotype father

M = phenotype mother

G_F = genotype father

G_M = genotype mother

A_F = additive genotype father

A_M = additive genotype mother

$$C_1 = \frac{\hat{V}_A + V_D}{V_P}$$

$$C_2 = \frac{\hat{V}_A}{\hat{V}_A + V_D}$$

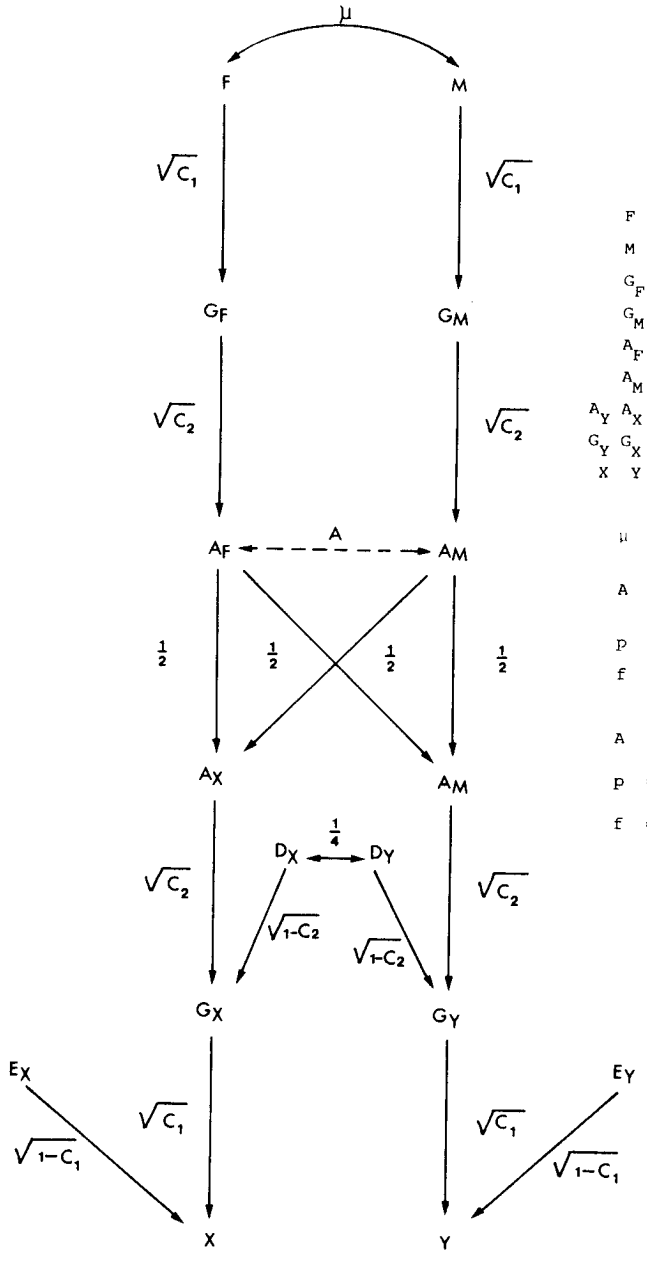
$\hat{V}_A$ = additive genetic variance, including assortative mating variance

V_D = variance due to dominance

V_P = phenotypic variance

Figure 1. Path diagram for Fisher's marital correlation due to assortative mating

are presented in Figures 1 and 3. Given μ, p, and f, C_1, C_2, and A can be determined, but one cannot, as is suggested by Jensen, predict μ, p, and f without making arbitrary assumptions about dominance, assortative mating, and the effect of the environment. In Jensen's paper (1972, p. 124) it is not made clear that the degree of assortative mating and dominance assumed by Burt and Howard is either chosen arbitrarily or derived from the empirically obtained correlations in order to maximize the fit between the model and the observations. The values of C_1, C_2, and A do not follow from Fisher's theory, but are model



F phenotype father
M phenotype mother
G_F genotype father
G_M genotype mother
A_F additive genotype father
A_M additive genotype mother
A_Y A_X additive genotype child
G_Y G_X genotype child
X Y phenotype child

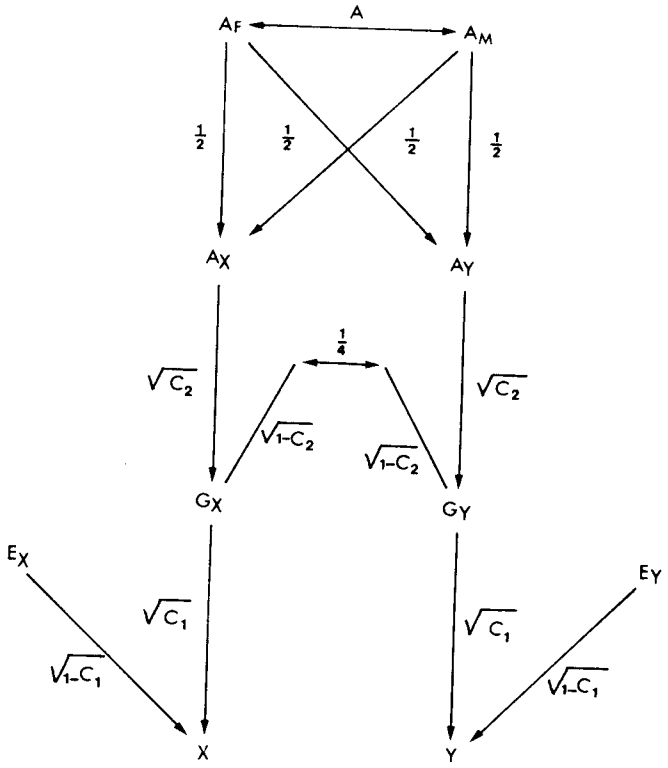
μ marital correlation of phenotypes
A marital correlation of additive genotypes
p parental correlation
f fraternal correlation

$A = C_1C_2\mu$

$p = \frac{1}{2}C_1C_2(1+\mu)$

$f = \frac{1}{4}C_1(1+C_2(1+2A))$

Figure 2. Path diagram for Fisher's complete model



$$r_{XY} = p = \sqrt{C_1}\sqrt{1-C_2}\ \tfrac{1}{4}\sqrt{1-C_2}\sqrt{C_1} \qquad \tfrac{1}{4}C_1 - \tfrac{1}{4}C_1C_2$$
$$+\ 2\sqrt{C_1}\sqrt{C_2}\ \tfrac{1}{2}.\tfrac{1}{2}\sqrt{C_1}\sqrt{C_2} = +\ \tfrac{1}{2}C_1C_2 =$$
$$+\ 2\sqrt{C_1}\sqrt{C_2}\ \tfrac{1}{2}\,A\,\tfrac{1}{2}\sqrt{C_2}\sqrt{C_1} \quad +\ \tfrac{1}{2}C_1C_2\,A$$

$$p = \tfrac{1}{4}C_1\,(1 - C_2 + 2C_2 + 2C_2\,A\,)$$
$$= \tfrac{1}{4}C_1\,(1 + C_2\,(1 + 2A)).$$

Figure 3a. Path diagram for Fisher's fraternal correlation under assortative mating

parameters to be estimated on the basis of observed correlations. Depending upon the degree of dominance, assortative mating, and environmental influences one assumes Fisher's polygenetic model can predict any values for p and f between 0 and 1. The (second) set of theoretical values mentioned by Jensen and Eysenck for intelligence data, for example, are based upon the analysis by Burt and Howard, who have used precisely Fisher's polygenetic model and assumed

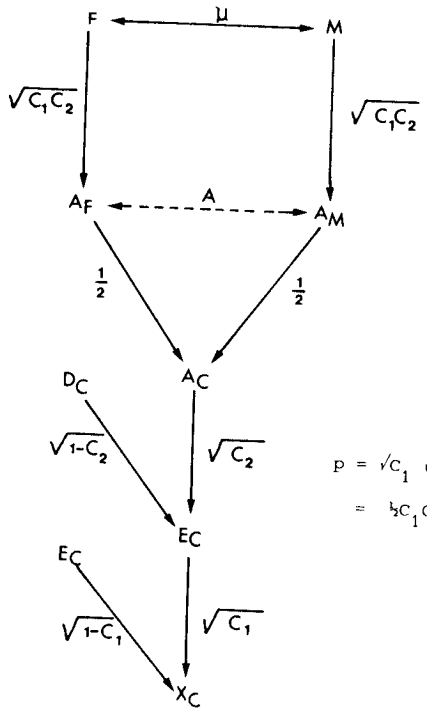


Figure 3b. Parental correlation under assortative mating

$C_1 = .95$ and $C_2 = .75$. Jinks and Fulker (1970) have shown that these values for C_1 and C_2 are almost the same as the values they obtain by using a weighted least-squares method in fitting the model. This is not very surprising, of course, because Burt and Howard 'predicted' the parental and fraternal correlations quite accurately. In order words Burt and Howard fitted the polygenetic model quite well in one lucky trial, but they did not make any quantitative prediction. One might just as well have reversed the values C_1 and C_2 in the Burt and Howard analyses, assuming virtually no dominance and considerable environmental influence and still produce a fairly accurate fit.

All this is not to say that Fisher's model does not make any predictions at all. It does predict a particular rank order of kinship correlations, but so does any model which is based upon parallel similarities between individuals in various degrees of kinship. The model does predict a regression effect, but so does any model which predicts an imperfect correlation. The model does predict inbreeding regression depending upon particular gene frequencies and directional dominance, but only a purely environmental model would have problems in explaining such an effect. This criticism of Fisher's model should not be misunderstood. We only want to make clear that the model developed

by Fisher does not make precise quantitative predictions of kinship correlations but, as a model, can only be fitted more or less successfully to observed correlations. Other models have been developed and can be constructed which do the job just as well and, as we will show later on, sometimes even better, depending upon the underlying assumptions. The fact that some of these well-fitting models are based upon unrealistic assumptions merely indicates that a good fit is not the only criterion in deciding between various models. Without further development of theory on the nature of genetic and environmental influences, a strong hereditarian or environmental position must be regarded as premature.

SEWALL WRIGHT'S ALTERNATIVE APPROACH: PATH ANALYSIS

Eaves (Eaves *et al.*, 1977) has recently argued that environmental critics of biometrical genetics have so far not advanced alternative quantitative models for the explanation of individual differences. By and large this is true, but that does not mean that the views of critics like Anastasi (1958), Hogben (1933), Woolf (1952), and Lippman (1923) are unimportant because they were not expressed in a quantitative form. The fact that their ideas have recently been taken up by geneticists like Rao *et al.* (1974), Cavalli-Sforza and Feldman (1973), Plomin *et al.* (1977), and Eaves himself and developed quantitatively indicates the contrary. Moreover, it should be pointed out that an alternative quantitative approach to behaviour genetics has been around for exactly as long a time as Fisher's approach, but has been ignored almost completely by Fisher and those who followed in Fisher's footsteps. Almost at the same time as Fisher wrote his paper on the correlation between relatives on the supposition of Mendelian inheritance, Wright (1921) introduced path analysis in a series of articles for exactly the same purpose. Although the two approaches are formally similar as we will show, the translation of the two models has never been made (see, however, Li, 1968), and one of the aims of this paper is to indicate how the two approaches relate to each other. The formal similarity of two approaches does not imply, however, that there are no substantial differences between the two research traditions. From the very beginning Wright points out that 'the characteristics of an individual may be looked upon as determined by two classes of factors—those which are internal and those which are external', or environmental in a broad sense. In Wright's analysis the environmental factors are separated into two components,—the common tangible environment and the intangible factors which are not common between individuals. In Fisher's original publication no such distinction is made and Fisher even points out that variance which is due to non-inheritable causes can also originate from irregularities of inheritance. Only many years later did Burt devote special attention to the common environ-

ment is Fisher's model. The second important difference between Fisher and Wright can be found in their treatment of dominance. Whereas dominance plays a central role in Fisher's analysis, Wright regards dominance as an important difficulty for his method of path analysis and proceeds by showing that the grading of factors, with and without taking dominance into account, are correlated. It would take us too far to discuss to detail the differences between Fisher and Wright in this respect, but from an example given by Wright on the inheritance of the piebald pattern in guinea-pigs it is clear that what is called dominance by Fisher is treated as common environment by Wright, because

$$4e^2 = (1 - C_2)C_1$$

where e is the common environment as defined by Wright, and C_1 and C_2 are as previously defined. Since

$$C_1 = \frac{V_A + V_D}{V_P} \quad \text{and} \quad C_2 = \frac{V_A}{V_A + V_D}$$

it follows that

$$4e^2 = \frac{V_A + V_D}{V_P} - \frac{V_A}{V_A + V_D} \cdot \frac{V_A + V_D}{V_P} = \frac{V_D}{V_P}$$

It is interesting to note that one of the earliest, more sophisticated, path analyses of data on human intelligence was made by Wright as early as 1931. Using data from one of the best studies on adopted children by Burks (1928), Wright develops several path models of which the most interesting feature is that he takes into account the correlation between genetic and environmental factors. He also assumes that the phenotype of the parents and their environment influences the environment of the child, which in turn influences the phenotype of the child. The study of Burks is usually referred to as a study in which it is shown that home environment explains no more than 10–20 per cent of the phenotypical variance in IQ, but in Wright's analysis environmental influences are also included in a miscellaneous factor, together with non-additive effects of genes and gene–environment interaction effects. Heritability is therefore defined in the narrow sense and shown to be lower for parents than for children. Although the possibility that h^2 could be different for different age populations is in principle recognized in the biometrical tradition, because h^2 is a population parameter, this fact has only recently received attention.

The main reason for this oversight, as we see it, is that the more detailed analysis of environmental factors suggested by environmental critics in the

1930s and 1940s and carried out to some extent by Wright was never followed up. In psychology the work by Burt is the prototype of the research tradition based upon Fisher's important article of 1918. It was Burt who explained Fisher for psychologists, and it is to Burt that an author such as Jensen unfortunately has paid his respects. It is for this reason that we will have to turn now briefly to the work by Burt and Jensen.

BURT'S TWINS AND JENSEN'S REVIVIFICATION OF THE NATURE–NURTURE CONTROVERSY

Burt is probably best known for his (in)famous study of monozygotic twins reared apart (MZA), but we would underestimate the importance of his work if we pay attention only to this study, which, moreover, has been criticized recently because of the probably unjustified suspicion of fraud. Much more important is the paper by Burt and Howard to which we have referred above. In discussions of Burt's work it is often overlooked that for Burt intelligence was *by definition* completely genetically determined because it denotes that part of the general cognitive factor which is attributable to the individual's genetic constitution. Only when one keeps this in mind does it make sense that Burt did not want to rely exlusively on mental tests, but wanted in addition, by retesting and interview methods, to allow for the influence of an exceptionally favourable or unfavourable cultural environment. Burt wanted an 'environment-free' measure of intelligence which again was important to him from a social point of view, because he believed that social conditions should not prevent gifted children from developing their talents. Kamin's attempt (Kamin, 1974) to ridicule Burt on these grounds is therefore quite misplaced. The fact that Burt took the trouble to measure intelligence 'environment-free' indicates already that he was well aware of the positive correlation between genetic and environmental factors in the case of intelligence and school achievement. Burt therefore attached great importance to the study of twins and of orphans. To differentiate systematic and random environmental effects Burt and Howard introduced the data on monozygotic twins reared apart, arguing that the difference between the heritability estimate based on Fisher's analysis and the direct estimate based on the correlation of the monozygotic twins reared apart reflects systematic environmental effects, including genetic–environment. However, this is only true if one assumes that the environments of MZA are uncorrelated and that there is no gene–environment correlation in that case. The fact that the socio-economic status of the foster-parents and the natural parents were uncorrelated in Burt's study is of course not sufficient evidence for making this assumption. Rearing children apart does not affect for example the gene–environment correlation caused by an effect of genes on the environment.

In fact there is some gene–environment covariation in Burt's study for the

unadjusted IQ data, but Burt has removed this source of variance, as we have seen, in his final assessments. The limitations of the Fisher–Burt model become quite clear when one applies the model to the average correlations published in the literature. First of all h^2 based on marital, parental, and fraternal correlations appears to be >1.00; which is the same problem Fisher had in analysing Pearson's data on stature. If we solve the problem in the same way Burt does, we find that there is a significant amount of gene–environment covariation. Using a 'parallelogram formula' Burt shows for his own data that there is a positive correlation ($r=.234$) between heredity and environment for the unadjusted test scores, a correlation which disappears in the final assessments. It seems to us that the heart of the matter of human behaviour genetics is, indeed, as Lewontin (1975) has pointed out again and again, that one cannot separate experimentally heredity and environment as orthogonal independent variables. One may test for between-family gene–environment covariation as Jinks and Fulker (1970) do and include a covariance component in the model, depending upon the outcome of such a test, but an alternative approach certainly is to include gene–environment correlations and environmental covariation from the start in a more general model. This is precisely what we have done in the second part of this contribution. By developing, moreover, all possible models within those constraints it is possible to show the consequences of the specific assumptions made by Fisher, Burt, and Jinks and Fulker.

Before we present these analyses we should perhaps say a few words about Jensen's contribution to behaviour genetics and the IQ debate. There is very little new information in Jensen's 1969 review article. It is, apart from its controversial implications for educational intervention, a selective and misleading summary of previous studies. Perhaps the most deceptive part of the paper is the crucial table of kinship correlations taken from Erlenmeyer-Kimling and Jarvik (1963). These correlations are, Jensen writes 'based on a wide variety of tests, administered under a variety of conditions' (Jensen, 1969, p. 48). According to Jensen the compatibility of a polygenetic hypothesis can be appreciated by comparing the median values of the obtained correlations with the theoretical values. What cannot be appreciated is, of course, the fact that there is considerable variation between studies for the same kinship relations. The original publication by Erlenmeyer-Kimling and Jarvik clearly shows this variation, as do comparable reviews by Fuller and Thompson (1960) and Jencks *et al.* (1972). In another publication Jensen (1972) has shown more sensitivity for this problem, because he reports as limits of h^2, based on twin studies, extreme values of .42 and .93, with some values exceeding 1.00. In this respect it is also curious that Jensen clearly recognizes that within- and between-family gene–environment covariation may occur, but nowhere does he take this into account in his calculations of h^2.

Jinks and Fulker, Jencks, Eaves *et al.*, Rao *et al.*, and Cavalli-Sforza and

Feldman have carried the biometrical and path analytical approach beyond the elementary level of Jensen's analysis, but as we will show in the next section, this does not imply that the work by the Birmingham school, especially, is free from criticism.

A GENERAL MODEL AND A REANALYSIS OF SOME STUDIES

In this section we explain the models of 'psychometric genetics', and we reanalyse some of the prime examples. We discuss the usual assumptions in some detail, and we compare the conclusions that can actually be drawn from an analysis of this sort with the conclusions that are usually drawn.

1. **Families**

We study populations defined by the following five symmetric relationships.

(1) MZT: monozygotic twins reared together by their own parents.
(2) MZA: monozygotic twins reared apart, i.e. in different foster-families.
(3) FST: full sibs (or dizygotic twins) reared together by their own parents.
(4) FSA: full sibs (or dizygotic twins) reared in different foster-families.
(5) URT: unrelated children reared together in the same foster-family.

2. **Assumptions**

2.1 *Additivity*

A person x has an (observed) phenotype score P_x, which is an additive combination of his (unobserved) genotype score G_x, his (unobserved) environment score E_x, and an error component R_x. In symbols

$$P_x = G_x + E_x + R_x. \tag{1}$$

Thus we assume that both genotypes and environments can be represented as one-dimensional scales. Moreover, we assume that environment is static, and finally we assume additivity. All these assumptions are extremely unrealistic on the level of the individual. They are consequently used in a descriptive sense, to describe or transform some properties of populations.

2.2 *Comparability*

In each of the five populations we are studying, the random variables G, E, R have the same means, variances, and covariances. The error component R is

uncorrelated with both G and E. We define

$$\mathrm{E}(G^2)=g^2 \tag{2a}$$

$$\mathrm{E}(E^2)=e^2 \tag{2b}$$

$$\mathrm{E}(R^2)=r^2 \tag{2c}$$

$$\mathrm{E}(GE)=\alpha ge \tag{2d}$$

By using standardized variables G, E, R we can now rewrite (1) as

$$P_x=gG_x+eE_x+rR_x. \tag{3}$$

The comparability assumption can be tested by comparing means and variances of P_x in all five populations. It can also be criticized on theoretical grounds: twinning is, to some extent, heritable; twins have very special prenatal and postnatal environments; adopted children come from special environments and are reared in special environments; covariances between genotypes and environments are different in adopted and natural children; birth order and family size effects are ignored; age is ignored.

2.3 *Correlations*

2.3.1 $\mathrm{E}(G_xG_y)$

The correlation between genotypes of MZ is unity. The correlation between genotypes of FS is $0\leqq\mu\leqq 1$. The correlation between genotypes of URT is $0\leqq\eta\leqq\mu\leqq 1$.

2.3.2 $\mathrm{E}(E_xE_y)$

The correlation between environments of persons reared together is unity. The correlation between environments of persons reared apart is $0\leqq\pi\leqq 1$.

2.3.3 $\mathrm{E}(E_xG_y)=\mathrm{E}(E_yG_x)$

For MZT, MZA, FST, and URT we have $\mathrm{E}(E_xG_y)=\alpha$ (this follows from our previous assumptions). For FSA we have $0\leqq\mathrm{E}(E_xG_y)=\beta\leqq\alpha$.

Again many criticisms are possible on theoretical grounds. MZ twins are not identical genetically, environments of persons reared together are not identical, environments of twins are more alike than those of sibs, and environments of MZ twins are more alike than those of DZ twins.

3. **Path diagrams**

The assumptions we have made so far can also be presented in the form of path diagrams. They are drawn in Figure 4a–e. The covariances between

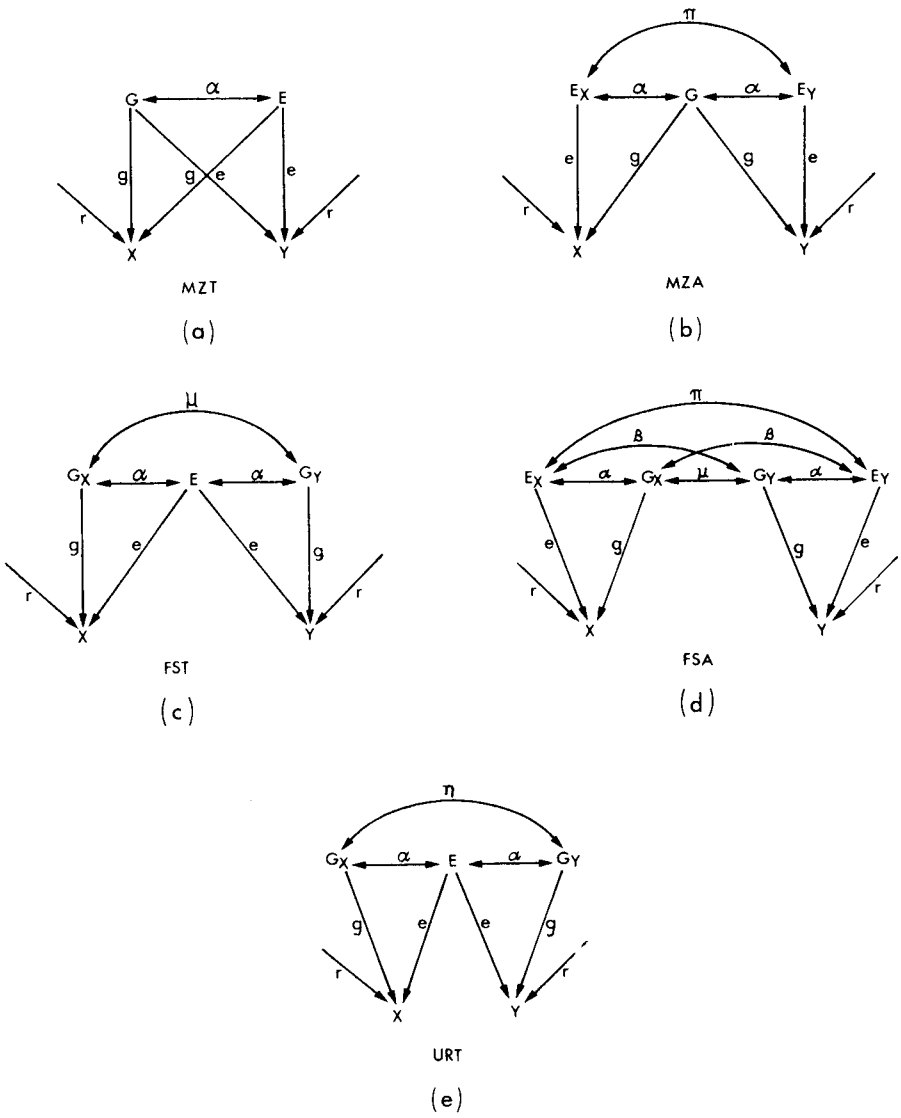


Figure 4. Path diagram for MZT, MZA, FST, FSA, and URT

phenotypes can be read directly from the diagrams.

$$\text{COV MZT} = g^2 + e^2 + 2\alpha ge \tag{4a}$$

$$\text{COV MZA} = g^2 + \pi e^2 + 2\alpha ge \tag{4b}$$

$$\text{COV FST} = \mu g^2 + e^2 + 2\alpha ge \tag{4c}$$

$$\text{COV FSA} = \mu g^2 + \pi e^2 + 2\beta ge \tag{4d}$$

$$\text{COV URT} = \eta g^2 + e^2 + 2\alpha ge \tag{4e}$$

$$\text{VAR} = g^2 + e^2 + 2\alpha ge + r^2 \tag{4f}$$

4. Linear model formulation

The model presented in (4) has eight parameters and only six observables. Consequently the parameters are not identified, and we must rewrite the model in terms of six identifiable parameters. We also want these parameters to be non-negative, and we want them to vary independently. Thus we cannot choose ηg^2, μg^2, and g^2 as parameters, because they are connected by the restrictions $\eta g^2 \leqq \mu g^2 \leqq g^2$. Consequently, we choose ηg^2, $(\mu - \eta)g^2$, and $(1-\mu)g^2$. For the environmental and covariance components we proceed in the same way. This gives the following parametrization of (4), in which $\mathbf{c}$ is the vector with the five different covariances and the common variance.

$$\mathbf{c} = \mathbf{A}\theta. \tag{5}$$

The 6×6 non-singular design matrix A is given by

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
MZT	1	1	1	1	1	0
MZA	1	1	0	1	1	0
FST	0	1	1	1	1	0
FSA	0	1	0	0	1	0
URT	0	0	1	1	1	0
VAR	1	1	1	1	1	1

(6)

An alternative ordinal characterization of this parametrization will be given in the section 5. The new parameters are obviously identifiable (because $\mathbf{A}$ is non-singular). They are related to the path diagram parameters in the following way:

$$\theta_1 = (1-\mu)g^2 \tag{7a}$$

$$\theta_2 = (\mu - \eta)g^2 \tag{7b}$$

$$\theta_3 = (1-\pi)e^2 \tag{7c}$$

$$\theta_4 = 2(\alpha - \beta)ge \tag{7d}$$

$$\theta_5 = \eta g^2 + \pi e^2 + 2\beta ge \tag{7e}$$

$$\theta_6 = r^2 \tag{7f}$$

If we analyse intraclass correlations we only use the leading 5×5 submatrix of **A**.

5. **Ordinal interpretation**

The model (5), with the additional restriction $\theta \geqq 0$, is of course equivalent to the system:

$$\mathbf{A}^{-1}\mathbf{c} \geqq 0. \tag{8}$$

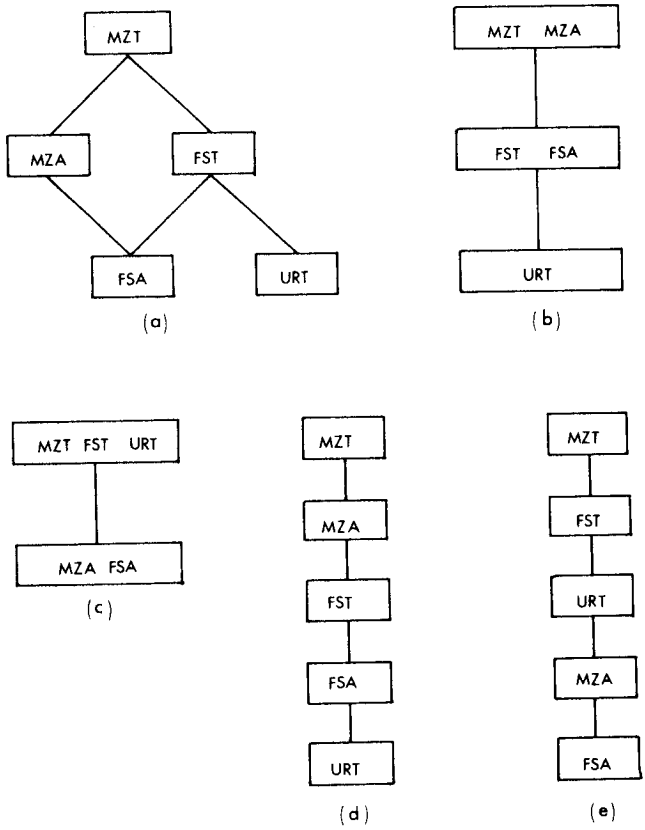


Figure 5. All reasonable models: (a) any reasonable model, (b) a purely genetic model, (c) a purely environmental model, (d) a lexico-genetic model, and (e) a lexico-environmental model

The inverse of **A** is the matrix:

	MZT	MZA	FST	FSA	URT	VAR
θ_1	+1	0	−1	0	0	0
θ_2	0	0	+1	0	−1	0
θ_3	+1	−1	0	0	0	0
θ_4	−1	+1	+1	−1	0	0
θ_5	0	0	−1	+1	+1	0
θ_6	−1	0	0	0	0	+1

(9)

Consequently, we can estimate the components of θ by the simple formula:

$$\hat{\theta} = \mathbf{A}^{-1}\mathbf{c}. \tag{10}$$

An important point is that we can interpret the elements of θ in terms of the parameters of the path model, using equations (7). For this interpretation we have to make all the assumptions of section 2, and these assumptions are essentially untestable because (5) is a saturated model. Alternatively we can use the inequality interpretation (8), (9) more directly, without translating back to the model (1). The ordinal interpretation can be extended. In Figure 5a we have drawn the partial order which both environmentalists and hereditarians would consider reasonable. A linear model is called reasonable if any non-negative linear combination of the model vectors satisfies this partial order. The model vectors satisfy this partial order. The ordinal constraints define a polyhedral convex cone, which can be defined dually as the set of all non-negative linear combinations of its edges. There are straightforward algorithms to find the set of all edges of a given cone. In our particular case it follows that all reasonable models can be described by the following eight model vectors.

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
MZT	1	1	1	1	1	1	1	1
MZA	0	1	0	1	0	1	1	1
FST	0	0	1	1	1	1	1	1
FSA	0	0	0	0	0	0	1	1
URT	0	0	0	0	1	1	0	1

(11)

It follows that our path model (4) is reasonable (at least in this very specific sense), because (6) shows that it uses model vectors 2, 5, 6, 7, 8 from (11). In fact our path model is equivalent to the reasonable partial order in Figure 5a,

together with the additional ordinal restrictions

$$\text{FSA} \geqq (\text{FST} - \text{URT}), \tag{12a}$$

$$(\text{FST} - \text{FSA}) \geqq (\text{MZT} - \text{MZA}). \tag{12b}$$

Other examples of 'reasonable' models are given in Figures 5b–e.

The Jinks and Fulker model

Jinks and Fulker (1970) have studied the following model in some detail.

	G_1	G_2	E_1	E_2
MZT	1	1	0	1
MZA	1	1	0	0
FST	0	1	0	1
FSA	0	1	0	0
URT	0	0	0	1
VAR	1	1	1	1

(13)

We have chosen the names also used by Jinks and Fulker for their components. By comparing (6) and (13) we see that their G_1 is our θ_1, their G_2 is our θ_2, their E_1 is our θ_6, and their E_2 is our θ_3. The interpretation of the components suggested by Jinks and Fulker is as follows. The first component G_1 is the within-family genetic variance, due to genetic differences of siblings. The second component G_2 is the between-family genetic variance, due to genetic similarities of siblings. The third component E_1 is the within-family environmental variance, due to differences in treatment of persons reared together. And the fourth component E_2 is the between-family environmental variance, due to similarities in treatment of persons reared together. By comparing (13) and (6) we see that Jinks and Fulker make all the assumptions we made, and two additional ones. The first one is random placement, which implies in this context that $\eta = \pi = \beta = 0$. The second one is no gene–environment covariance, which implies $\alpha = 0$. Alternatively, this can also be written as $\theta_4 = \theta_5 = 0$, which is equivalent with

$$\text{FSA} = \text{FST} - \text{URT}, \tag{14a}$$

$$\text{FST} - \text{FSA} = \text{MZT} - \text{MZA}. \tag{14b}$$

Thus (11) reduces to (6) if we impose the restrictions (12), and (11) reduces to (13) if we impose the stronger restrictions (14).

Equations (14) can be interpreted by using the simple factorial design

	T	A
MZ	MZT	MZA
FS	FST	FSA
UR	URT	URA

(15)

If we assume that URA = 0, then (14) implies that (15) is additive. Thus the Jinks and Fulker model can be interpreted as MZ ≧ FS ≧ UR and T ≧ A and additivity. Or to put it differently, the Jinks and Fulker model is the only reasonable model (in the sense of section 5) that makes (15) additivie. It is clear that this interpretation of the model is a 'fixed effects' version of the 'random effects' model (1).

Burt's assessments

For our first example we reanalyse some correlations given by Burt (1966). The same data have been analysed by Jinks and Fulker (1970), and Fulker (1974), using the Jinks and Fulker model of formula (12). The main purpose of our reanalysis is to show that it is extremely misleading to study a single submodel of the saturated model. Another interesting result is that the Jinks and Fulker model, with the Jinks and Fulker interpretation, gives a very high heritability estimate. The data are:

	Correlation	Sample size
MZT	.92	95
MZA	.87	53
FST	.53	264
FSA	.44	151
URT	.27	136

We have analysed—without VAR and consequently θ_6—the data by fitting all $2^5 - 1$ submodels of the saturated model (6), using an iterative programme based on z-transformed intraclass correlations. The programme is called INTRA, it is written in APL, and it minimizes the transformed chi square by using Gauss–Newton iterations. As the initial estimate we use weighted least squares on the correlations. These estimates are usually reported by the Birmingham school, but they are not very satisfactory from a statistical point of view, because the normal and chi-square approximations are usually very poor with these sample sizes. Our results are given in Table 2. We do not give standard errors of the estimates. Because the estimates have large intercorrelations the standard errors can be quite misleading. The significance of a component can

Table 2 Reanalysis of Burt's assessments[a]

	Model	χ^2	θ_1	θ_2	θ_3	θ_4	θ_5
1	00001	130.298	—	—	—	—	.6029
2	00010	154.696	—	—	—	.6405	—
3	00100	228.144	—	—	.6029	—	—
4	01000	111.830	—	.6632	—	—	—
5	10000	138.510	.9060	—	—	—	—
6	00011	120.886	—	—	—	.1978	.4427
7	00101	130.298	—	—	.0000	—	.6029
8	01001	101.244	—	.3897	—	—	.2734
9	10001	10.746	.4570	—	—	—	.4491
10	00110	135.311	—	—	−.2694	.8723	
11	01010	79.948	—	.4459	—	.2781	—
12	10010	44.542	.4546	—	—	.4514	—
13	01100	101.498	—	.5615	.1455	—	—
14	10100	72.444	.5774	—	.3541	—	—
15	11000	13.983	.4057	.5003	—	—	—
16	11100	5.915	.3933	.4392	.0937	—	—
17	11010	4.692	.3492	.3771	—	.1797	—
18	10110	42.438	.4695	—	.0485	.4029	—
19	01110	69.300	—	.4342	−.1770	.4381	
20	11001	3.397	.4057	.2269	—	—	.2734
21	10101	8.844	.4594	—	.0415	—	.4184
22	01101	98.054	—	.4204	.0909	—	.1825
23	10011	10.733	.4546	—	—	.0088	.4427
24	01011	79.942	—	.4510	—	.2817	−.0084
25	00111	101.502	—	—	−.2694	.4296	.4426
26	11110	2.588	.3640	.3771	.0485	.1312	—
27	11101	.212	.4013	.2473	.0561	—	.2174
28	11011	2.104	.3747	.2579	—	.0887	.1847
29	10111	8.629	.4694	—	.0485	−.0397	.4427
30	01111	69.257	—	.4204	−.1785	.4296	.0222
31	11111	—	.39	.26	.05	.04	.18

[a]Negative estimates are underlined

be read from the table by looking at models which do not have that particular component.

The number of degrees of freedom for the chi-squares in Table 2 can be computed quite simply by counting the zeros in the model specification. Thus model 30 has 1 df, model 9 has 3 df, and so on. We can also perform all kinds of hierarchical analyses by using Table 2, and all kinds of partitionings of the total chi-square. If we want to test if θ_5 is significant, we first have to specify the model which we assume to be true. If we assume the saturated model 31 to be true, we test $\theta_5 = 0$ by subtracting the chi-square of model 31 from that of model 26. This gives 2.588, with one degree of freedom. If we assume model 20 to be true, then we can test $\theta_5 = 0$ by subtracting the chi-square of 20 from that

Table 3a Burt's assessments

Model	χ^2	θ_1	θ_2	θ_3	θ_4
0001	120.886	—	—	—	.6405
0010	194.335	—	—	.6029	—
0100	90.528	—	.7244	—	—
1000	104.701	.9060	—	—	—
0011	101.502	—	—	−.2694	.8723
0101	79.942	—	.4510	—	.2734
1001	10.733	.4546	—	—	.4514
0110	90.222	—	.7577	−.0354	—
1010	38.635	.5774	—	.3541	—
1100	12.691	.3747	.5314	—	—
1110	5.910	.3950	.4361	.0953	—
1101	2.104	.3747	.2579	—	.2734
1011	8.629	.4994	—	.0485	.4029
0111	69.257	—	.4204	−.1785	.4519
1111	—	.39	.26	.05	.22

Table 3b Jencks' data

Model	χ^2	θ_1	θ_2	θ_3	θ_4
0001	119.827	—	—	—	.5912
0010	135.516	—	—	.5899	—
0100	139.069	—	.6138	—	—
1000	953.189	.9478	—	—	—
0011	118.016	—	—	−.1720	.7616
0101	97.458	—	.2321	—	.3817
1001	32.946	.3791	—	—	.5688
0110	105.668	—	.2738	.3411	—
1010	23.048	.4050	—	.5663	—
1100	57.240	.3577	.5902	—	—
1110	1.630	.3810	.2320	.3582	—
1101	15.629	.3577	.2085	—	.3817
1011	17.318	.4018	—	.2090	.3598
0111	96.024	—	.2306	−.1493	.5310
1111	—	.38	.21	.22	.16

of 15. This gives 10.586, again with one degree of freedom. Thus θ_5 is not significant in the saturated model, but it is significant in model 20.

A complete analysis of Table 2 along these lines would be very long and very uninteresting. We can translate all models that fit reasonably well back to the path model. Thus model 15 must be interpreted as $e^2=\eta=0$, model 20 can be interpreted as $e^2=0$ or as $\pi=1$ and $\alpha=\beta$, model 17 does not make sense because $\theta_3=\theta_5=0$ implies $e^2=0$, which implies $\theta_4=0$. Model 21 implies $\mu=\eta$ and $\alpha=\beta$, which is still not enough to identify g^2 and e^2. If we assume $\mu=0$ in

21, then $g^2 = .4594$. If we assume $\pi = 0$ and $\alpha = 0$ then $g^2 = .8778$. And so on. We can summarize the conclusions as follows.

(1) The Burt data are consistent with the following estimates: $.45 \leqq g^2 \leqq .90$, $.00 \leqq e^2 \leqq .40$, and $.00 \leqq 2\alpha ge \leqq .30$.
(2) The Burt data are also consistent with the hypothesis that $\eta = \mu$ and/or that $\pi = 1$. More generally, some of the submodels that cannot be rejected do not make any sense at all (if interpreted in the path model we have used).
(3) If we compare P-values, then there are five models with a better fit than the Jinks–Fulker model. All these models give a lower estimate of $G_1 + G_2$, which is .83 for the Jinks–Fulker model, and between .65 and .75 for the 'better' models.
(4) If we assume random placement, then $.73 \leqq g^2 \leqq .90$, $.00 \leqq e^2 \leqq .10$, $.00 \leqq 2\alpha ge \leqq .18$.

We could go on discussing the consequences of $\pi = 1$, and/or $\mu = 0$, both of which are not contradicted by the data, but the point is probably clear. If one works within the path model only very imprecise conclusions are possible, some thoroughly silly models cannot be rejected, and many different interpretations are possible, even within a single analysis, and even within the path model (depending on the 'extra' assumptions). We have also analysed the Burt data by leaving out MZA, by leaving out URT, and by leaving out all twins. The results of the corresponding hierarchical analyses make the ranges of the possible genetic, environmental, and covariance contributions even wider. If we leave out FSA we can compare Burt's data with those of Jencks (1972). The results are presented in Table 3. If we leave out FSA it becomes impossible to distinguish θ_4 and θ_5. We collapse them to a new θ_4 with interpretation $\theta_4 = \eta g^2 + \pi e^2 + 2\bar{\alpha} ge$. Jinks and Eaves (1974) have argued that the results of the two studies are quite similar.

Our table shows that they are quite different. In the first place the Jencks data illustrate the familiar goodness-of-fit problem: if the sample sizes are large enough, almost everything is rejected. The only submodel that cannot be rejected has $g^2 = .51$, $e^2 = .36$, $r^2 = .13$. For the Burt data this model gives $g^2 = .83$, $e^2 = .10$, $r^2 = .07$. There seems to be some difference between white American Caucasians and white English LCC Caucasians. For the MZT–MZA–DZT data of Newman *et al.* (1937) we find $.35 \leqq g^2 \leqq .88$, $.00 \leqq e^2 \leqq .56$ for Stanford IQ, $.41 \leqq g^2 \leqq .89$ and $.00 \leqq c^2 \leqq .52$ for Otis IQ, and $.08 \leqq g^2 \leqq .53$, $.40 \leqq e^2 \leqq .88$ for Stanford Achievement.

Australian twins

The second example we analyse in some detail is Martin (1975). He analysed school achievement data for a representative sample of Australian twins,

whose zygosity was carefully determined. The path model is

	θ_1	θ_2	θ_3
MZ COV	1	1	0
DZ COV	0	1	0
MZ VAR	1	1	1
DZ VAR	1	1	1

(16)

and the interpretation of the parameters is

$$\theta_1 = (1-\mu)g^2 \tag{17a}$$

$$\theta_2 = \mu g^2 + e^2 + 2\alpha ge \tag{17b}$$

$$\theta_3 = r^2 \tag{17c}$$

Martin uses a slightly different model. If we assume that there is no linkage and epistasis, and we use the model of Fisher (1918), then

$$\mu = \tfrac{1}{2}p_A + \tfrac{1}{4}p_D \tag{18}$$

with p_A the proportion of genetic variance that is additive, and $p_D = 1 - p_A$ the residual, due to dominance deviations. This suggest the reparametrization

	ε_1	ε_2	ε_3	ε_4
MZ COV	1	1	1	0
DZ COV	$\frac{1}{2}$	$\frac{1}{4}$	1	0
MZ VAR	1	1	1	1
DZ VAR	1	1	1	1

(19)

in which the interpretation of the parameters is

$$\varepsilon_1 = \sigma_A^2 \tag{20a}$$

$$\varepsilon_2 = \sigma_D^2 \tag{20b}$$

$$\varepsilon_3 = e^2 + 2\alpha ge \tag{20c}$$

$$\varepsilon_4 = r^2 \tag{20d}$$

Models (16) and (19) are linearly equivalent, they span the same subspace of dimension three. Model (16) has a very simple ordinal interpretation. It simply predicts that $\mathrm{COV}_{MZT} \geqq \mathrm{COV}_{DZT}$ and that $\mathrm{VAR}_{MZT} = \mathrm{VAR}_{DZT}$. Martin uses submodels (1001) and (0011) of (19). The first model (the 'genetic' one) predicts

that $COV_{MZT} = 2COV_{DZT}$, the second one (the 'environmental' one) predicts that $COV_{MZT} = COV_{DZT}$. From (17) we also see that the twin method is biased, the hypotheses that $e^2 = 0$ can never be rejected by data.

Again, we have fitted a hierarchy of models by an iterative programme based on the so-called Wilson–Hilferty transformation of the between- and within-MZ and DZ variances. The program is called BETWIT, it is written in APL, it uses the usual estimates as starting points and improves them by Gauss–Newton iterations. The cube root transformation of the variances is needed to make the normal and chi-square approximations any good. The results are given in Table 4a (for the complete model (19)) and Table 4b (for the submodel with $g^2 = 0$). It is clear that for seven of the eleven subjects the hypothesis $g^2 = 0$ cannot be rejected. For the remaining subjects (English, geography, maths 1, IQ) the data are consistent with heritabilities between .45 and 1.00. If we assume that $e^2 = 0$ in the complete model the heritabilities are all around .80 (except for history), but the model gives estimates of μ larger than .60 or smaller than .40 in six out of eleven cases. If we assume that $\mu = \frac{1}{2}$ in the complete model then the heritabilities for the four subjects mentioned above are larger than .75. But in all four cases the estimate of e^2 is negative. It seems that the only conclusion that can be drawn from an analysis of these data is that for four out of eleven subjects COV_{MZT} is considerably larger than COV_{DZT}.

We illustrate the analysis of within and between mean squares with a final example, due to Adams *et al.* (1976). Again this is a representative sample of twins, all of the same age, with zygosity diagnosed by the familiar 'two peas in a pot' question. Results for a verbal test are given in Table 5a, results for a non-verbal test in Table 5b. For both tests $g^2 = 0$ cannot be rejected, assuming $e^2 = 0$ gives $g^2 = .70$ for the verbal and $g^2 = .76$ for the non-verbal test, but also $\mu = .84$ for the verbal and $\mu = .76$ for the non-verbal test. We have also analysed

Table 4 Martin's data

	θ_1	θ_2	θ_3	χ_1^2	θ_2	θ_3	χ_2^2
English	79.02	75.11	41.28	.446	109.68	88.09	17.297
French	165.55	199.24	74.35	1.326	299.62	159.12	5.850
History	46.39	65.76	131.49	1.108	80.37	161.11	1.880
Geography	112.76	70.60	42.96	.036	119.24	108.60	11.177
Maths 1	150.69	123.95	63.12	2.359	204.78	146.54	15.471
Maths 2	69.93	193.15	64.84	1.136	229.32	104.32	4.681
Physics	44.84	188.97	71.73	1.168	217.25	93.47	2.087
Chemistry	90.25	330.02	48.10	.092	377.47	94.23	4.083
Science 1	32.85	185.02	73.57	1.349	203.15	91.73	2.041
Science 2	94.45	182.14	83.42	1.878	237.30	133.90	4.710
IQ	58.15	46.24	26.84	1.218	72.34	54.25	9.076

(a) Complete model (model 19).

(b) Submodel $g^2 = 0$.

Table 5 Data of Adams *et al.*

Model	χ^2	θ_1	θ_2	θ_3	χ^2	θ_1	θ_2	θ_3
111	.000	4.796	24.434	12.415	.000	5.268	18.604	7.455
101	39.405	29.332	—	12.422	40.533	23.952	—	7.458
011	1.512	—	25.859	15.776	3.880	—	20.174	11.148

(a) Path model, verbal test.

(b) Path model, non-verbal test.

Model	χ^2	ε_1	ε_2	ε_3	ε_4	χ^2	ε_1	ε_2	ε_3	ε_4
1101	.000	68.505	−39.276	—	12.415	.000	50.543	−26.671	—	7.455
1011	.000	9.592	—	19.638	12.415	.000	10.536	—	13.336	7.455
0111	.000	—	6.394	22.835	12.415	.000	—	7.024	16.848	7.455
1001	8.489	29.353	—	—	10.336	7.453	23.279	—	—	6.406
0101	22.089	—	29.144	—	10.826	21.810	—	23.209	—	6.642
0011	1.512	—	—	25.859	15.776	3.880	—	—	20.174	11.148

(c) Martin's model, verbal test.

(d) Martin's model, non-verbal test.

the data by using model (22), the results are in Tables 5c and 5d. Because of the linear dependencies the first three submodels give the same fit (either all genetic variance is additive or non-additive, heritability is around .20–.30). Model (1001) gives heritablities of .74 for the verbal test and .78 for the non-verbal test, model (0011) shows that $g^2=0$ gives an even better fit. The only sensible conclusion is that for these data $COV_{MZT}=COV_{DZT}$ cannot be rejected.

Conclusions

(1) All tests and estimates only make sense within the particular model we have adopted. This model is, at least partly, based on assumptions which are both untestable and implausible. We can test additional specifications within a saturated model, but the saturated model itself remains untestable. More complicated models (which include parents, grandparents, uncles) need more implausible assumptions (the number of assumptions grows exponentially with the number of relationships).

(2) Selection of a single submodel from the saturated model can be very misleading. A hierarchical analysis shows that the data are usually consistent with a very wide range of models and estimates.

(3) The interpretation of the model must be separated from the model itself. Many different interpretations of a simple linear model are usually possible, of which the gene–environment path model is only a very special one. The model itself only predicts a partial order of the correlation coefficients, which can usually be explained both genetically and environmentally. A non-significant chi-square means, according to standard statistical practice, that we cannot reject the model. But even if we cannot reject the model empirically we can reject the interpretation on *a-priori* grounds.

The quantification of the social environment

As was shown in the previous section, the biometric model, used by Jinks and Fulker, assumes no gene–environment covariation or correlated environments. In terms of the path analyses presented, the Jinks and Fulker model assumes $\eta=\pi=\beta=0$ and $\alpha=0$. In all cases where a better fit was obtained some of these paramters had positive values, and consequently our estimates of h^2 were lower for these better fitting models. This means that in the Jinks and Fulker analysis the environment is treated as error variance and for the rest is equated with growing up in the same family. In this respect biometrical genetics had not progressed far beyond the notions already advanced by Galton. Until quite recently the quantification of the environmental components has been rarely attempted in biometrical genetics. Although Galton realized, as we have seen, that the environment of the child is influenced by the phenotype of the

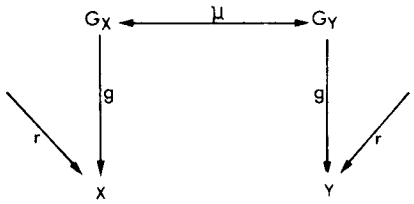


Figure 6. Path diagram for Fisher's simple model

mother and the genotype of the child itself, he did not take this into account in his analysis. Fisher explicitly treated the environment as a random factor. A path diagram for the Fisher model would look like the one in Figure 6. Fisher's specific contribution was to derive (from 'Mendelian' assumptions) expressions for the various genetic correlations μ in terms of three more basic parameters. In the case of random mating with no epistasis and no linkage we need only two parameters, the proportion of variance due to additive gene action and the proportion due to dominance deviations. If mating is assortative we need an additional parameter to quantify the degree of assortative mating. It is possible to draw a path diagram which gives the same expressions as Fisher's model for the parent–offspring, the sib–sib, and the marital correlation. Such a diagram looks complicated and not very informative. In fact even Wright does not use path analysis if he discusses Fisher's 'analysis of variance' model.

From Figure 6 we see that there is no common environment affecting phenotypes of both X and Y in Fisher's model. This is introduced later by Mather and Jinks (1971), Burt (1966), Jinks and Fulker (1970), but still without assuming a correlation between *G* and *E*. It was Wright (1931) who presented the first quantitative analysis in which gene–environment covariation was assumed, based upon the hypothesized effect of the parents' environment and phenotype on the child's environment as can be seen in Figure 7.

It is this model which has been extended by Jencks (1972), Rao *et al.* (1974), Cavalli–Sforza and Feldman (1973), Plomin *et al.*, (1977), and Eaves *et al.* (1977). It would be impossible to discuss here, in detail, these various models in which the environmental component is further specified. Their common features are, however, readily illustrated in Wright's path diagram. First of all there are, in addition to the genetic path from the parental genotype to the child's genotype, two cultural paths from the parental phenotype and the parental environment, to the environment of the child. If we think of the latter primarily in terms of the social interactions in which the child is involved, these paths indicate that the educational behaviour of the parents is determined by their phenotype and their own (social) environment or background. It is this part of the environment which has been called 'passive' by Plomin *et al.* and 'cultural transmission' by Eaves *et al.* However, the social

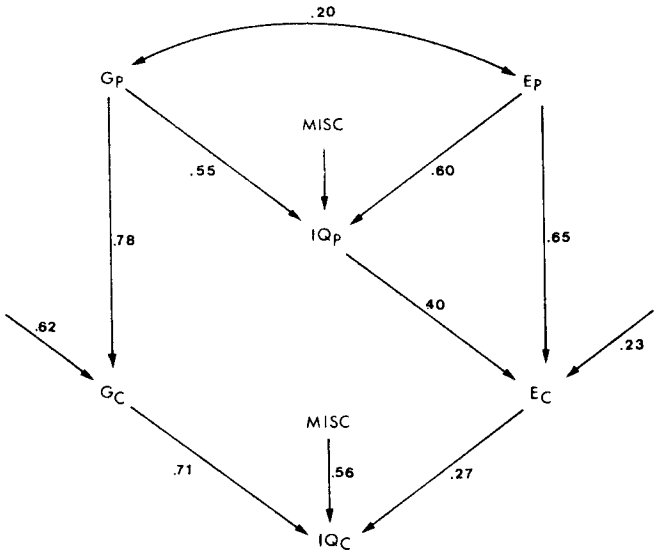


Figure 7. Path diagram for Wright's analysis of Burk's data

environment of the child does not only consist of parental behaviour, but in general it consists of the behaviour of all the people with whom it interacts. For obvious reasons attention has been paid in the first place to sibling effects, i.e. the effect of the phenotype of other children in the family on the environment of the child, but other children (friends) or adults (teachers) might be included as well. In the third place several authors have allowed for a direct effect of the genotype on the environment of the child. Eaves *et al.* and Jinks and Fulker see this effect primarily as a selection of the environment by the genotype, whereas Jencks and Plomin *et al.* also include the possibility of the environment reacting differently to different genotypes. Eaves *et al.* and Jinks and Fulker have argued that in a single culture at a single point in time, this effect cannot be separated from 'direct' effects of the genes. This is only true, however, if one assumes that the child's environment cannot be measured directly. Moreover, in our own analysis this effect, together with the passive environment effect and/or cultural transmission, is included in the covariance component and not in the genetic component.

In our analysis so far we have, however, not included direct sibling effects. It seems reasonable to assume that such effects are absent when siblings are reared apart, but when siblings are raised in the same family a realistic model would have to include such influences. Because we have assumed that the environments of children reared in the same family is identical, sibling effects imply at the same time a reciprocal effect of the child's phenotype on its environment. Such reciprocal effects have never been considered in biometrical

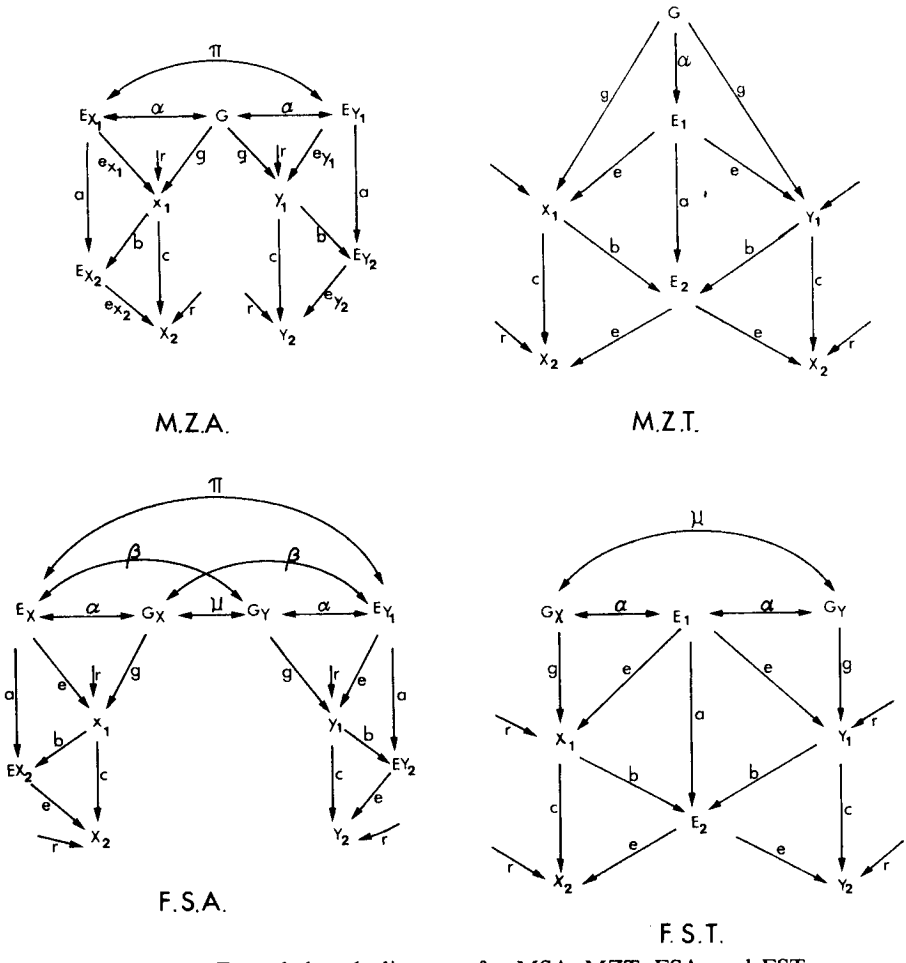


Figure 8. Extended path diagrams for MSA, MZT, FSA, and FST

genetics and are difficult to deal with in path analysis. A fairly simple solution, suggested by Wright (1960), is to extend the path analysis in time. Applied to behaviour genetics this would imply repeated measurement of phenotypes and environments. In Figure 8 we have presented the minimal extension for the cases of MZA, MZT, FST, and FSA.

If we define the environment in terms of social behaviour of parents and siblings such an extension does not imply the introduction of new latent variables, and the new path coefficients can be given a direct empirical interpretation as the stability of the phenotype and the environment. It is clear that the diagrams can easily be extended to families with more than two children, measured at time points t_1, $t_2, \ldots,$ t_n. An even more interesting

extension would be to use continuous time, and to describe the path model by differential equations. Of course the quality of the IQ data, which cannot even be used to test the appropriateness of simple static models such as those in Figure 4, is so poor that testing these dynamic models is virtually unthinkable.

The detailed implications of the path models in Figure 8 are not studied in this paper. In terms of the covariances at times t_1 and t_2 some of the more interesting conclusions are.

$$\begin{aligned} \mathrm{MZA}_1 &\geqq \mathrm{MZA}_2 \\ \mathrm{MZA}_1 - \mathrm{FSA}_1 &\geqq \mathrm{MZA}_2 - \mathrm{FSA}_2 \end{aligned}$$

If there is no covariance and random placement then MZA estimates g^2, and MZA − FSA estimates $(1-\mu)g^2$. As a consequence static models will find a heritability estimate which decreases with age. Although there is indeed a tendency of this kind in the published IQ data for MZA these results should not be considered as evidence, because all kinds of tests were used and because the Burt data are suspect. In fact at this moment hardly any direct evidence is available in the literature which could provide a test of a reciprocal model. However, there are some studies which provide indirect evidence which is in agreement with the hypotheses formulated above.

First of all Wright and Rao *et al.* also arrive at the conclusion that heritability estimates of IQ are lower for adults than for children, although they present completely different evidence. One could, moreover, systematize available estimates in the literature according to the age of the subjects involved. This has not yet been done. The best approach would be, of course, through longitudinal studies. Unfortunately, as far as is known, such studies are not available. There are several longitudinal studies on the development of IQ, but these studies do not contain information on the family relationships of the subjects. The only study we were able to find is a study by McCall *et al.* (1973) in which some longitudinal IQ data are reported in graphical form of monozygotic and dizygotic twins and triplets raised together. Inspection of these data shows that with age the differences between children become larger and that a conventional heritability estimate becomes lower with increasing age. Of more interest is perhaps the analysis performed by McCall *et al.* on the total group of children. It appears that for the age range from $2\frac{1}{2}$ to 17 years there is a considerable intra-individual variation of IQ. This variation is not random but related to the behaviour of the parents, which in turn is related to the phenotypes of the children. The result of this process is exactly what we would expect in the case of positive reciprocal effect, i.e. initial differences become larger and McCall *et al.* are able to show that at least for the younger age period different divergent development curves can be distinguished.

From a social–psychological point of view the reciprocal effects described above can be seen as a process of positive feedback or positive reinforcement

in social interaction. Thus, in a general sense there is nothing new in the idea that an individual's behaviour or dispositions can be changed through social interaction and that an individual's personality affects the social interactions in which he takes part. However, studies of social interactions have so far not yet been related to the study of population differences in IQ, scholastic achievement, etc. We believe that the time has now come to integrate development social interaction studies with the study of behaviour genetics. Studies on social interaction and cognitive development in the tradition of Piaget, and studies of language development and parent–infant interaction point the way (Dunn, 1976). Moreover, several interesting mathematical models have recently been developed for the analysis of parent–infant interaction. For example, Thomas and Martin (1976) have described a continuous state and a discrete state model in which not only the effect of the parent on the child, but also the reverse effect is taken into account. The confluence model developed by Zajonc and Markus (1975) to describe differences in intelligence depending upon family size and birth order also takes into account the effect of the child upon its environment. The difficulty with the models of Thomas and Martin and Zajonc and Markus is that from the point of view of behaviour genetics they are as yet incomplete models. Zajonc and Markus relate directly phenotypes of family members to each other, whereas Thomas and Martin relate parent and infant behaviour to each other. This suggests that we should include in our model the child's behaviour in addition to its phenotype, because it is through its behaviour that the child affects its social environment and is affected by it.

The extension of behaviour genetics in the direction of social interaction processes does not, of course, deny the importance of genetic factors for the description of population differences. Rousseau already recognized this two centuries ago when he wrote that 'l'inégalité naturelle doit augmenter dans l'espèce humain par l'inégalité d'institution'.

REFERENCES

Adams, B., Ghodsian, M., and Richardson, K. Evidence for a low upper limit of hereditability of mental test performance in a national sample of twins. *Nature*, 1976, **263**, 314–316.

Anastasi, A. Heredity, environment, and the question 'How?'. *Psychological Review*, 1958, **65**, 197–208.

Burks, B. S. The relative influence of nature and nurture upon mental development: A comparative study of foster parent–foster child resemblance and true parent–true child resemblance. *27th Yearbook of the National Society for the Study of Education*, 1928, **27**(1), 219–316.

Burt, C. The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *British Journal of Psychology*, 1966, **57**, 137–153.

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

Cavalli-Sforza, L. L. and Feldman, M. Cultural versus biological inheritance phenotypic transmission from parents to children (a theory of the effect of parental phenotypes on children's phenotypes). *American Journal of Human Genetics*, 1973, **25**, 618–637.

Dunn, J. How far do early differences in mother–child relations affect later development. In P. P. E. Bateson and R. A. Hinde (Eds.),*Growing points in ethology*, Cambridge: Cambridge University Press, 1976.

Eaves, L. J. Assortative mating and intelligence: an analysis of pedigree data. *Heredity*, 1973, **30**, 199–210.

Eaves, L. J., Last, K. Martin, N. G., and Jinks, J. L. A progressive approach to non-additivity and genotype–environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology*, 1977, **30**, 1–42.

Elderton, E. M. A summary of the present position with regard to inheritance of intelligence. *Biometrika*, 1923, **14**, 378–408.

Erlenmeyer-Kimling, L. and Jarvik, L. F. Genetics and intelligence: a review. *Science*, 1963, **143**, 1477–1479.

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

Frogatt, P. and Nevin, N. C. The 'law of ancestral heredity' and the Mendelian–ancestral controversy in England, 1889–1906. *Journal of Medical Genetics*, 1971, **8**, 1–36.

Fulker, D. W. Applications of biometrical genetics to human behaviour. In J. H. F. van Abeelen (Ed.), *The genetics of behaviour*. Amsterdam: North-Holland, 1974.

Fuller, J. L. and Thompson, W. R. *Behavior genetics*. New York: Wiley, 1960.

Galton, F. Hereditary talent and character. *Macmillan's Magazine*, 1865, **12**, 157–166, 318–327.

Galton, F. *Natural inheritance*, London: Macmillan, 1889.

Galton, F. *Hereditary Genius. An inquiry into its laws and consequences*. London: Macmillan, 1892 (2nd ed.).

Heymans, G. *Gesammelte kleinere Schriften*. Teil III, Den Haag: Nijhof, 1927.

Hogben, L. *Nature and nurture*. London: Williams and Norgate, 1933.

Jencks, C. *et al. Inequality*. Harmondsworth: Penguin Books, 1972.

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

Jensen, A. R. *Genetics and education*. London: Methuen, 1972.

Jinks, J. L. and Fulker, D. W. Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychological Bulletin*, 1970, **73**, 311–349.

Jinks, J. L. and Eaves, L. J. IQ and inequality. *Nature*, 1974, **248**, 287–289.

Kamin, L. J. *The science and politics of IQ*. New York: Wiley, 1974.

Lewontin, R. C. Genetic aspects of intelligence. *American Review of Genetics*, 1975, **9**, 387–405.

Li, C. C. Fisher, Wright and path coefficients. *Biometrics*, 1968, **24**, 471–483.

Lippmann, W. Mr. Burt and the intelligence test. *New Republic*, 1923, **34**, 263–264.

Martin, N. G. The inheritance of scholastic abilities in a sample of twins. II, Genetical analysis of examination results. *Annual of Human Genetics*, 1975, **39**, 219–229.

McCall, R. B. Appelbaum, M. I., and Hogarty, P. S. Developmental changes in mental performance. *Monthly of the Society for Research in Child Development*, 1973, **38**, 15–69.

Mather, K. and Jinks, J. L. *Biometrical genetics*. London: Chapman and Hall, 1971.

Newman, H. H., Freeman, F. N., and Holzinger, K. J. *Twins: A study of heredity and environment*. Chicago: University of Chicago Press, 1937.

Pearson, K. On the law of ancestral heredity. *Proceedings of the Royal Society*, 1898, **62**, 386–412.

Pearson, K. On the laws of inheritance in man II. *Biometrika*, 1904, **3**, 131–190.

Plomin, R. De Fries, J. C., and Loehlin, J. C. Genotype–environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, 1977, **84**, 309–322.

Rao, D. C. Morton, N. E., and Yee, S. Analysis of family resemblance. II, A linear model for familial correlation. *American Journal of Human Genetics*, 1974, **26**, 331–359.

Rousseau, J. J. Discourse sur l'inégalité parmi les hommes. In B. Gagnebin and M. Raymond (Eds), *Oeuvres complètes*. Paris: Gallimard, 1959–1964.

Thomas, E. A. C. and Martin, J. A. Analysis of parent–infant interaction.*Psychological Review*, 1976, **83**, 141–156.

Woolf, B. Environmental effects in quantitative inheritance. In E. C. R. Reeve and C. H. Waddington (Eds.), *Quantitative inheritance*. London: H.M.S.O., 1952.

Wright, S. Systems of matings. *Genetics*, 1921, **6**, 111–178.

Wright, S. Statistical methods in biology. *Proceedings of the American Statistical Association*, 1931, **26**, 155–163.

Wright, S. The treatment of reciprocal interaction with or without lag in path analysis. *Biometrics*, 1960, **16**, 423–445.

Zajonc, R. B. and Markus, G. B. Birth order and intellectual development. *Psychological Review*, 1975, **82**, 74–88.