Br. J. math. statist. Psychol. (1977), **30**, 1–42 *Printed in Great Britain*

A progressive approach to non-additivity and genotype–environmental covariance in the analysis of human differences.

L. J. Eaves, Krystyna Last, N. G. Martin and J. L. Jinks

No aspect of human behaviour genetics has caused more confusion and generated more obscurantism than the analysis and interpretation of the various types of non-additivity and non-independence of gene and environmental action and interaction—genotype–environment interaction and covariation, dominance and assortative mating. A comprehensive framework of theory and method is outlined in which these and other contributions to individual differences can be critically assessed.

1. Introduction

There is a view, widely held and frequently expressed, that genotype–environmental interaction ($G \times E$) and genotype–environmental covariation (Cov GE) are factors which preclude a worthwhile analysis of individual differences in human populations. Such is the view expressed in a variety of ways, for example, by Moran (1973), Layzer (1974), Lewontin (1974), Feldman & Lewontin (1975), and others. Whilst these authors allude to the possible importance of $G \times E$ and Cov GE they offer no adequate specification of these effects, nor do they suggest how such effects may be detected.

There is, however, a large body of data and theory relating to the analysis of $G \times E$ in species other than man (e.g. Haldane, 1946; Mather & Jones, 1958; Bucio-Alanis *et al.* 1969; Jinks & Perkins, 1970; Jinks & Connolly, 1975; Mather & Caligari, 1975), there have long been attempts—not always successful—to specify Cov GE in man (e.g. Cattell, 1960; Loehlin, 1965), and a much-publicized paper by Jinks & Fulker (1970) deals with the principles and pitfalls in the analysis of both $G \times E$ and Cov GE in man. In spite of a substantial literature in this area, there is still considerable ignorance about the theoretical specification of these effects, their practical analysis and their biological significance. In this paper we seek to clarify many areas of misunderstanding surrounding all three aspects of the analysis of human differences.

2. 'Assumptions' and scaling tests

The obvious needs to be stated. There are two types of assumption: those which can be tested, given adequate data, and those which cannot. An assumption in biometrical genetics is meant to be tested. The analytical power of biometrical genetics is due to the fact that assumptions are made not merely for the convenience of estimating parameters but in order that the model they imply might be tested. This fact is not always appreciated by critics who regard an assumption as a mark of weakness rather than as a necessity for any attempt to test null hypotheses. A study of individual differences which tests no assumptions about the causes of variation is little more than an exercise in the juggling of numbers. By contrast, the statistical tests employed in biometrical genetics, the so-called 'scaling tests' (see Mather & Jinks, 1971, for a recent account), however simple or elaborate they may be, are developed precisely to test one or more assumptions about the origin of differences. Much of our paper is

devoted to a detailed examination of the methods by which particular assumptions may be tested and our confidence that such assumptions are adequately tested in practice.

3. Non-additivity

3.1. *Classification*

In our attempts to analyse variation we have to consider two types of non-additive effect which may contribute to individual differences, namely, genetical non-additivity and genotype–environmental interaction.

(*a*) *Genetical non-additivity*. Continued directional selection, acting on a trait, is expected to increase the relative importance of non-additive genetical effects, particularly directional dominance and epistasis of the duplicate-gene type (e.g. Mather, 1943, 1967, 1973; Breese & Mather, 1960; Broadhurst & Jinks, 1974).

There exist two fundamental misconceptions relating to the analysis and biological significance of non-additive gene action. One is typified by Feldman & Lewontin (1974) who claim, on the basis of a purely mathematical argument with no experimental foundation, that selection should eradicate additive variation almost entirely whilst leaving a great surfeit of dominance for traits related to fitness. They have misunderstood the basis of the argument which rests not on algebra but experiment. The trait is unknown to quantitative genetics for which non-additive variation exists without a substantial additive component. Numerous experimental studies suggest that genetical non-additivity is expressed against a background of additive gene action for traits related to fitness.

The second misunderstanding is illustrated by Morton (1974) who asserts: 'The notion of dominance deviations for polygenes seems far fetched'. That this is untrue will be attested by any practical quantitative geneticist with experience of the importance of dominance and epistasis in organisms other than man (e.g. Comstock & Robinson, 1952; Jinks, 1955; Breese & Mather, 1957, 1960; Jinks & Jones, 1958; Kempthorne, 1960; Jinks & Broadhurst, 1963; Hanson & Robinson, 1963; Jinks & Perkins, 1969, 1970*a*).

At the very simplest level the superior vigour of commercial F_1 and double cross hybrids and the ubiquity of inbreeding depression in outbreeding species are commonplace manifestations of non-additive gene action. Certainly Morton, advocating path coefficients as an analytical device, is faced with a conceptual difficulty in the analysis of any kind of non-additive variation but it is mistaken to ignore such aspects of proven significance in other organisms merely because their analysis in man is more difficult. Indeed, as Morton and his associates refine their approach in an attempt to include non-additive effects the method should become very similar to that of biometrical genetics and ought to give the same answer with adequate data.

(*b*) *Genotype–environmental interaction*. Whilst some authoritative writers have attempted to justify the assumption of genetical additivity, perhaps in order to simplify the processes of specifying models and estimating their parameters, several others have lent their support to another view that $G \times E$ is by contrast a widespread and substantial component of individual differences. At the same time, it is suggested, $G \times E$ effects in man are both substantial on the one hand and undetected on the other, with the implication that attempts to estimate and interpret other population parameters are thereby seriously in error.

Animal and plant experiments certainly demonstrate that $G \times E$ is widespread whenever a set of genotypes is grown in a variety of controlled or even uncontrolled

environments. In recent years, there has been considerable success in the detection, analysis and interpretation of $G \times E$. It is dangerous, however, to exaggerate the general significance of $G \times E$. If variation in man has any similarity to variation in other organisms we would conclude that a trait was atypical if more than about 20 per cent of the measured variation could be attributed to $G \times E$. An effect of this magnitude would be important but not overwhelming. We shall consider the feasibility of detecting such effects in man.

3.2. *Systematic and non-systematic effects*

Before we can attempt to analyse non-additive effects, much less presume to comment on their significance, we must recognize the practical distinction between:

(a) systematic ('directional' or 'scalar') effects; and

(b) unsystematic ('ambidirectional') effects.

This distinction is important in the consideration of genetical non-additivity and genotype–environmental interaction for both practical and theoretical reasons.

At the level of practical analysis both kinds of effect contribute to variation and so we can expect that they will contribute to second-degree statistics, i.e. to variances and covariances of raw observations. Thus, for example, all genes showing dominance will tend to increase variability and so contribute to the non-additive component of second-degree statistics. On the other hand, first- and third-degree statistics will only be affected by non-additive effects if there is a net directional component. Thus, only directional dominance (as opposed to ambidirectional dominance) will result in inbreeding depression or in skewness within segregating families. A similar distinction is essential with respect to $G \times E$. Although all kinds of $G \times E$ contribute to second-degree statistics, detailed analysis of $G \times E$ in experimental organisms suggests that part, often a substantial part, of the $G \times E$ variation is due to the effects of genes which are associated in effect or distribution with the genes which contribute to the differences between genotypes averaged over environments (e.g. Perkins & Jinks, 1968, 1973; Fripp & Caten, 1973; Mather & Caligari, 1975). Such associations can often be broken by artificial selection (Perkins & Jinks, 1968, 1973; Brumpton, 1973; Jinks & Connolly, 1975), which implies that they may well have been maintained solely by natural selection. The very fact that $G \times E$ exists at all means that sensitivity or reaction to the environment is itself under genetical control, and may consequently be altered by selection just as selection may change or stabilize any other aspect of the phenotype which is under genetical control.

Both directional non-additive genetical effects, among which we include directional dominance, and systematic $G \times E$ interaction of the type we have discussed may contribute to third-degree statistics. Jinks & Fulker adapted the methods used for the detection of such $G \times E$ to the detection of systematic $G \times E$ in man. We shall see that only a relatively small part of the environmental variation need be due to genes associated systematically with overall genetical differences in order to be detected by their approach. The practical importance of such systematic interactions is that they allow a measure of prediction about the relative efficiency of the same degree of environmental manipulation at different points of the scale of measurement.

3.3. *Non-additivity and scale*

In a recent book, Kamin (1974, p. 152) dismissed an analysis of genotype–environment interaction for a cognitive trait on the grounds that: 'Whether or not we observe an

interaction depends upon our choice of scale. The choice of scale is arbitrary—as is the advice provided to educators . . . .' In fact, of course, quantitative predictions of any kind are dependent on the choice of scale. This fact is rarely understood outside those areas of biology and psychology in which the attempt to make quantitative predictions flourishes, albeit prematurely in some instances.

Virtually all kinds of systematic non-additive effects could be removed by a change of scale, either by a simple transformation or by altering, in a psychometric study, the difficulty and discriminating power of the items which constitute a particular test. By altering the relative amounts of information at different points in the scale of individual differences the psychometrician is altering the relative weightings given to different gene substitutions and environmental circumstances. He is thus, whether he likes it or not, adjusting the epistatic and genotype–environmental interactions expressed in the trait he is measuring. A change of scale is a change of trait. We shall see that the same items, handled in slightly different ways, give radically different pictures for the causes of individual differences. As long, however, as a particular scale is used for measurement and prediction, we are forced to accept whatever complications may be required in order to explain the causes of variation on that scale. It is a fallacy to suppose that there is a 'true' scale. Any scale is arbitrary. There are merely scales which are more satisfactory for some purposes than others. It is also a fallacy to suppose that a scalar transformation to remove non-additivity is really only 'hiding' the interaction. Gene substitutions and environmental effects have to be measured by their effect on the phenotype at some level or other. Often there will be conflict about the choice of scale because it is not to be expected that genetical and psychological considerations will always coincide in suggesting a choice of scale. A scale which satisfies a criterion of genetical additivity may be subject to genotype–environmental interactions and vice versa. A trait which has desirable genetical properties may display a complex and undesirable relationship with some external criterion variable.

The existence of genetical and genotype–environmental interactions of a systematic variety may alert the psychometrician to areas in which further test development may take place. Marked directional non-additivity, for example, may indicate a threshold in the scale beyond which measurement is difficult, or has simply not been attempted. We shall consider the consequences of different forms of scaling later.

The so-called 'problem of scale' is only a problem for those whose scientific inquiry has never proceeded beyond the limitations of broad qualitative statements unsupported by measurements. The long-standing use of the term 'scaling test' to the variety of methods used in biometrical genetics for the detection of many kinds of non-additivity is a testimony to their dependence on the scale used. The criteria of scaling are many, and will depend on the principal theme of a particular inquiry. Few, if any scales will satisfy all criteria. A particular choice of scale will be vindicated by the successful predictions it facilitates. In a biometrical–genetical context the situation is summarized simply by Mather & Jinks (1971, p. 63) in the following way:

> The scales of the instruments which we employ in measuring our plants and animals are those which experience has shown to be convenient to us. We have no reason to suppose that they are specially appropriate to the representation of the characters of a living organism for the purposes of genetical analysis. Nor have we any reason to believe that a single scale can reflect equally the idiosyncrasies of all the genes affecting a single character. . . . The scale on which the measurements are expressed for the purposes of genetical analysis must therefore be reached by empirical means. Obviously it should be one which facilitates both the analysis of the data and the interpretation and use of the resulting statistics.

Lord & Novick (1968, p. 22) state a similar view in a psychometric context as follows: 'If a particular interval scale is shown empirically to provide the basis of an accurately predictive and usefully descriptive model, then it is a good scale and further theoretical developments might profitably be based on it. Thus measurement (or scaling) is a fundamental part of the process of theory construction.'

4. Non-independence

The second major class of assumptions which are characteristic of preliminary attempts to explain human differences relate to the independence of genetical and environmental effects. Gene effects may be correlated *inter se* as a result, for example, of linkage disequilibrium stemming from assortative mating. Genetical and environmental effects may be correlated for a variety of reasons, including cultural transmission and sibling effects.

4.1. *Assortative mating*

The literature on assortative mating has recently been revived because of several papers dealing with aspects of Fisher's (1918) treatment of assortative mating, for a long time the basis of most analyses of the human mating system. As with all theoretical debates it is difficult to decide always whether various criticisms point to errors of argument which lead to serious errors of interference, or whether they are merely differences in degree of approximation well beyond the resolution of most practical studies. In the final analysis, any theoretical considerations must stand the test of data. Wilson (1973), for example, has argued that assortative mating without selection is unrealistic and offers a generalization of Fisher's approach which can take account of the fact that extremes may find it more difficult to mate. The practical crux of her theoretical argument rests on whether or not the population of spouses is significantly unrepresentative of the population of genotypes from which they are drawn. In a large and detailed study such differences are almost certain to be detected. Vetta & Smith (1974), however, have cast some doubt upon the correctness of the mathematical argument advanced by Wilson. Until this issue is resolved any practical application of her model must wait. On the other hand, Vetta & Smith (1974) and Vetta (1976) have examined Fisher's argument in great detail and have generally satisfied themselves of its mathematical propriety, even if they and others have some reservations about its likely practical application. In a recent communication, Vetta (1976) has suggested that Fisher's expectation for the parent offspring correlation is slightly in error because assortative mating is expected to introduce some covariation between the dominance deviations of parents and offspring. He examines the data analysed in Fisher's original paper and confirms that Fisher's estimates of genetical parameters are in error in the second and third significant figures (personal communication). Such differences, whilst no doubt mathematically significant, do not come within the resolution of practical studies at the present time. In fact, wherever Fisher's model of assortative mating has been applied and tested (see e.g. Eaves, 1973, 1975), the fit to real data has been fairly good. This is not to claim that other models of assortative mating might not give as good, or even better, fit. But we should bear in mind that in the last analysis more extensive data, and not mathematical theory, are the only way of resolving the practical issues.

The problem of assortative mating illustrates a more general problem in quantitative genetics, particularly when its methods are applied to man, and that is the problem of

resolution. Although we may list the many factors which contribute to human variation, we are only likely to detect some factors if they account for a considerable, and sometimes a very considerable, proportion of the variation. We shall discover this as we consider our ability to detect various effects later in this paper. The extent to which we can detect a given effect will always depend on its nature, its magnitude and the experimental design. Some effects can be detected with great certainty, others with great unreliability. In our experience of the analysis of human data, it is possible to show that some effects which account for as much as 20 per cent of the total variance are beyond resolution in unfavourable circumstances.

At the present time, therefore, our aims have to be much more modest than those of correcting the second and third significant figures of our estimates of parameters. We have to concentrate on testing those assumptions which we know to be testable in principle and whose failure could lead to major errors of inference.

Thus, with reference to assortative mating, it is a trivial task to demonstrate a marital correlation for a trait. It is slightly more exacting to demonstrate that such a correlation is having the expected genetical consequences in human populations. Eaves (1973) showed how the genetical consequences of assortative mating could be detected in one population without reference to the marital correlation. Many studies of IQ have suggested that the linkage disequilibrium resulting from assortative mating is having a significant effect on the amount and distribution of genetical variation in human populations. Even at this level, however, we have to recognize that studies with dimensions of those usually performed will almost certainly be unable to detect the genetical consequences of assortative mating unless there is considerable additive genetical variation for the trait under consideration. Furthermore, although under such circumstances we may be satisfied that a theory such as Fisher's is adequate to explain the variation, we would be more dubious about excluding a variety of derivative theories which predict very similar consequences.

One aspect of assortative mating theory which could quite easily become the basis of empirical study is the precise relationship between the various kinds of correlation between spouses. Most practical treatments have assumed that the genotypic correlation between spouses is a pale reflection of the phenotypic correlation between spouses, and that the latter is properly estimated from the correlation of measurements made on a single occasion. When Eaves (1973) attempted to estimate the degree of assortative mating for IQ without reference to the marital correlation and then employed the genotypic correlation to predict the correlation between spouses, it was found that the predicted marital correlation was closer to the observed value after correction for unreliability than to the raw correlation between parental IQ scores. Although it would be a mistake to use such non-significant differences as any more than a starting point for discussion, it does illustrate the possibility that the human organism, integrating data over a wide range of occasions and encounters, may achieve a more accurate assessment (albeit unconscious) of the innate abilities of potential mates than is provided by explicit psychological tests.

Fisher himself expressed doubt about the exact relationship between the phenotypic and genotypic correlations between spouses and gave different sets of expectations for the correlations between relatives depending on the precise model which was assumed for assortative mating. Generally it has been assumed (and not disproved) that the genetical correlation between spouses is a secondary consequence of their phenotypic correlation.

Fisher's treatment of assortative mating, and indeed many of the alternative approaches, assumes that genetical and environmental factors are additive and

independent. It is not easy to see how damaging genotype–environment interaction would be for the specification of assortative mating using Fisher's approach. In so far as the genes affecting sensitivity to the environment are independent in effect and distribution with respect to those which determine overall differences for the trait on which assortative mating is based it is difficult to see how the approach can be seriously in error. We can see, however, that, perhaps as a result of assortative mating or selection, the genes responsible for stability may become associated systematically with those which determine a particular overall expression of the trait in question. Thus, for example, it is not beyond the bounds of possibility that a genetical system could have evolved in which the genes which increase intelligence are associated with genes which promote stability of gene expression in a wide range of environments. Systematic $G \times E$ interactions of this type, however, are precisely those which the approaches of biometrical genetics, exemplified in this context by the approach of Jinks & Fulker, are most able to detect. In practice, Fisher's model of assortative mating has not been used, and we doubt whether it should be used, when systematic $G \times E$ interactions are known to be present.

The independence of genetical and environmental factors is the basis of Fisher's treatment of assortative mating. There seems still to be some confusion about the precise content of the assumptions that Fisher made in deriving his expectations. Providing environmental factors remain independent of genetical differences it does not matter for Fisher's model whether or not the environmental deviations are correlated for individuals reared in the same family. Once the quality of the environment in which a family develops depends on the genotypes of the parents who provide the environment (i.e. in the presence of cultural transmission), then individuals' genotypic deviations are no longer distributed independently of their environmental differences and Fisher's model may not be applicable to individual differences in the presence of such genotype–environmental covariance.

4.2. *Genotype–environment covariance*

Cattell (1960) was one of the first authors to consider seriously the consequences of the possible covariation of genetical and environmental effects in human populations. The weakness of his approach, and virtually every subsequent approach to the problem, has been the lack of any precise theory for the origin of genotype–environment covariation (Cov GE), which can be cast in a quantitative form to allow any parsimonious and powerful treatment of variation in the presence of Cov GE. The specification of Cov GE is plagued by empiricism. As recently as 1975, Thomas maintained in connection with Cov GE: '. . . precisely how these terms may be viewed may be in dispute because substantive theory is lacking'. The fact that so many attempts to specify Cov GE have come to grief is because their authors have thought in statistical rather than biological terms. Their approach has been to write in a model virtually every conceivable covariance term involving genetical and environmental effects, and then to decide, by intuitive arguments, which genotype–environment correlations could be set to zero and which could be regarded as equal for the purposes of estimation. Often, as is the case most recently in Thomas' (1975) approach, there is no attempt at all to decide what restraints may operate upon the parameter values, with the result that quite arbitrary and misleading restraints are applied merely to obtain a solution. Indeed, Jinks & Eaves (1974) used the approach of specifying arbitrary restraints in order to solve for genotype–environmental correlation in an analysis of

IQ data. This approach is undesirable, and is a poor substitute for a theory which enables us to see quite clearly the relationships between genotype–environment covariance parameters in different kinds of statistics.

The classical approach is to specify Cov *GE* in terms of a genotype–environmental correlation (r_{ge}) and the genotypic and environmental standard deviations (σ_g and σ_e). This is the approach of Cattell (1960), Loehlin (1965), Jenks *et al.* (1973), Hogarth (1974), Jinks & Eaves (1974), Jensen (1975), Thomas (1975) and Goldberger & Lewontin (1976).

This approach is deficient. The relative magnitudes of r_{ge} in different kinds of family are determined by the source of environmental variation which covaries with genetical differences. For example, when the genotype–environmental correlation arises because one sibling forms a developmentally significant part of the environment of another, the environmental variation and the genotype–environmental covariation resulting from this can be specified precisely and economically in a way that leads to testable hypotheses. However, the usual approach, through the specification of r_{ge}, leads to an unnecessary multiplication of parameters because the genotype environment correlation and the environmental variance are interdependent in a way which depends on the degree of relationship.

In fact, it is possible, as Eaves (1976*a*, *b*) has shown, to parameterize genotype–environment covariance in a variety of ways consistent with meaningful biological and psychological theories. We may distinguish the following three kinds of Cov *GE* which can be specified in terms of a theoretical model. Two of these can be detected by quite simple studies.

(a) *Environments selected by genotypes.* In the event of superior genotypes seeking or establishing for themselves advantageous environments, that part of the environmental variation and the Cov *GE* which depends on genetical differences between individuals will be confounded with estimates of genetical variation. Any analysis of individual differences in a single culture, at a single point in time, will be unable to separate the 'direct' effects of the genes from those which operate through promoting selection of, or change in, the environment. Further, gene expression may be altered as a result of cultural change. A freer, more mobile society might display greater genetic variability because individuals are free to select the environment in which they develop. A restrictive more static society may reduce genetic variability in a variety of ways, simply by preventing genotypes from selecting or creating their own environments. Such changes in genetical variability resulting from the stimulation or suppression of one source of genotype–environment correlation may also be regarded, in another light, as a form of genotype–environment interaction. Because, in general, we are only able to assess the performance of any given array of genotypes within a single culture it may be difficult to design clear-cut demonstrations of this phenomenon.

(b) *Sibling effects.* Human beings often develop in the presence of siblings. Identical twins develop in the presence of a sibling of identical genotype; foster children may be raised in the presence of a sibling who is genetically unrelated. If the behaviour of siblings is important in development, either because siblings compete for available resources or cooperate in obtaining resources, then a characteristic pattern of Cov *GE* may emerge which can be clearly identified (Eaves, 1976*a*). Under these circumstances the genotype–environmental *correlations* are expected to show a complex functional relationship which is better represented by a simple model that recognizes those linear relationships expected to exist between the genotype–environmental *covariances*.

(c) *Cultural transmission.* Perhaps the most significant area of concern, at least theoretically, is the extent to which cultural transmission is an important component of individual differences. Cavalli-Sforza & Feldman (1973) have renewed interest in this area by offering an approach in which the effect of the phenotype of one individual (in this case a parent) influences environmentally the phenotype of another (in this case an offspring). Eaves (1976*b*) has examined several of the consequences of this approach for the analysis of randomly mating populations, and shows how covariance of genetical and environmental differences between families may be a consequence of cultural transmission perpetuating differences whose origin is ultimately genetical. As with the case of sibling effects, the effects of cultural transmission can be specified parsimoniously in terms of a mathematical model which suggests that cultural transmission has its own pattern of Cov *GE* for which diagnostic tests are simply devised.

The importance of genotype–environment covariance is twofold. Genotype–environmental covariance of the kinds we have distinguished is only possible if genetical influences are modifying the quality of the environment. The detection of Cov *GE* is thus important psychologically since it draws our attention to the personal aspects of an individual's environment rather than to the accidents of development. People become more significant in the environment than material things. Secondly, the detection of Cov *GE* may be important biologically because a population in which one genotype can affect the performance of another is a necessary prerequisite for any system of evolution by group or kin selection. It may be that Cov *GE* could provide some basis for deciding between traits on which natural selection operates on an individual basis, and those which may be subject to kin selection.

One form of Cov *GE* we have not considered is that which arises because of placement. There is inevitably an element of empiricism in the specification of such models, but perhaps even here theories about placement may be given some more concrete and testable form. Subsequently we shall consider in more detail the specification and detection of the various forms of genotype–environmental covariance. Although we consider only the specification of Cov *GE* for randomly mating populations and although an adequate theoretical specification for assortatively mating populations may still be elusive, we should note that the general consequences of the different kinds of Cov *GE* will persist whatever the mating system and for this reason we must not be misled into thinking that Cov *GE* is not detectable in the presence of assortative mating. In this respect the scaling tests we propose for the presence of Cov *GE* are scale free. In the presence of detectable Cov *GE*, however, especially when this is due to cultural effects (i.e. involving the covariance of genetical and environmental differences between families such as Jinks & Fulker suggested might be the case for educational attainments), we suspect that any further attempts to analyse gene action when there is assortative mating should be resisted until an adequate theoretical specification of the joint effects of mating system and culture has been realized.

5. Elementary considerations

When we are faced with the genetic analysis of any body of data several questions occur which form the basis of the subsequent analysis and interpretation:

(1) What are the appropriate statistics for summarizing the data prior to analysis?

(2) What set of assumptions is it proposed to test and how might they be formulated in terms of a quantitative model?

(3) What scaling tests can be devised to facilitate the testing of these assumptions and how might the parameters best be estimated?

(4) Given that a simple model fails how might it be extended?

(5) How powerful are these tests given the structure and size of the sample?

(6) How serious are the errors of inference which may follow from failure to disprove an assumption which is, in fact, false?

It may be thought that these questions are too basic to need repetition, but it is our view that more confusion stems from failure to consider these matters carefully than from any other. We now consider the first three issues, and defer examination of the other three to later sections.

5.1. *Choice of statistic*

For too long the correlation coefficient has been adopted as the starting point for an analysis of individual differences. Cattell's MAVA and the biometrical genetical approach are exceptions to what is virtually universal practice. Whatever the appeal of the intraclass correlation as a number, and however much we may be forced into using it because it is the statistic used in the past, we must recognize that simple-minded use of the correlation coefficient may well lead to the obscuring of highly significant features of the genetical and environmental system, to inefficient tests of others and to undetected biases in estimates of parameters. Correlations are only an effective starting point for an analysis of individual differences when the causes of individual differences are fairly simple. Otherwise information is wasted in the standardization of the different groups of data to unit variance. This is particularly misleading if hypotheses are to be tested whose failure is most clearly to be seen in a pattern of total variances. Thus, for example, one of the first signs of sex-linkage or sex-limitation may be a difference between the total variances of the two sexes. More important, however, is the expected pattern of total variances for different groups of individuals in the presence of genotype–environmental covariation. Such differences are better analysed than removed by a purely statistical device.

We can illustrate the problem by reference to a simple example. Consider the analysis of monozygotic twin data in the presence of covariation of genetical and environmental differences between families. Suppose we have data on twins reared together and twins reared apart in randomly chosen foster homes. Differences within pairs of twins reared apart will be due to within- and between-family environmental influences (E_1 and E_2 respectively). The component of variance between such pairs will reflect only genetical differences (G). For twins reared apart, there is expected to be no Cov GE. For twins reared together, variation within pairs will reflect simply environmental differences within families (E_1) whilst differences between pairs will reflect G, E_2 and, in the presence of genotype–environmental covariance, a contribution from this source (Cov $g_2 e_2$).

If we start with the raw data, and derive the mean squares of the analyses of variance within and between pairs for each twin type (i.e. producing four mean squares in all), we have four statistics and four parameters (G, E_1, E_2 and Cov $g_2 e_2$), as shown in Table 1. Thus, a perfect fit solution can be obtained to give estimates of all the parameters and to yield tests of significance of the parameters, although no further test of the model is possible and no further investigation of the genetical system can be undertaken. If we now standardize the data to give intraclass correlations, we have only two statistics, namely, the correlation of MZ twins reared apart and that for twins reared together. Although we can introduce one arbitrary constraint on the parameter values (e.g. that $G+E_1+E_2$ is unity), we still have three free parameters in our model with only two statistics available for their estimation. This means that, by

taking intraclass correlations, we have made it quite impossible to estimate the relative contributions of genetical and environmental effects and their covariation. By taking correlations information has been lost, in this instance to the point of making any solution impossible. This is an extreme example, but the general principle remains that standardization to unit variance reduces the power of certain crucial tests of assumptions which are comparatively powerful when the analysis of variance is used as a starting point for the analysis.

Table 1. Expectations of variance components for monozygotic twins reared together and apart (MZT and MZA), in the presence of covariance between genotypic and environmental differences between families

	G	E_1	E_2	Cov $g_2 e_2$
Between MZT	1	·	1	2
Within MZT	·	1	·	·
Between MZA	1	·	·	·
Within MZA	·	1	1	·

We have laboured this point because Morton (1974) has, for statistical rather than biological reasons, expressed the opinion that: 'The estimation theory should be developed in terms of the z transform of correlation, for which the normality assumption is less restrictive. Clearly, emphasis must be on tests of hypotheses rather than on estimation.' In fact, the approach to estimation advocated by Morton is that rejected by experienced quantitative geneticists precisely because it ignores many of the simple and powerful tests of hypotheses available with the raw statistics. At best there will be a loss of information; at worst parameters will not be estimable and any estimates which are obtained may be seriously biased (see Jinks & Fulker, 1970, pp. 324ff.). Whatever the practical and statistical convenience of the correlation coefficient, it is not an appropriate starting point for model fitting.

Instead, following the long experience of biometrical genetics, we start with the data summarized in terms of analyses of variance, identifying, for example, the mean squares within and between families; or, in cases where the analysis of variance might be inappropriate, we start with variances and covariances between relatives, for example, the variances and covariances of parents and offspring.

5.2. *Formulating a model for individual differences*

Anyone can write a model but there is little point in doing so unless the model embodies a testable null hypothesis. Often a simple model might be written which could fail in practice for a variety of reasons. Sometimes a model may fail but we may be unable to decide exactly what is causing the failure with the data available. Thus, for example, we may represent in a model the null hypothesis that the data give no reason to doubt the randomness of mating, the additivity of gene action, the absence of cultural effects and genotype–environment covariance, and the equality of environmental variances for different kinds of relatives. Such a simple model can be written and tested, even with data on identical and fraternal twins reared together. Any of the factors excluded in the null hypothesis might cause failure of the model, and although we might be able to find further data which suggest what is contributing to the inadequacy of the simple model, the twin data by themselves will be inadequate for this purpose. Thus, for

example, it is sometimes taken as self-evident that the environmental differences within MZ twins are smaller than those of DZ twins. If this is the case, then a simple model which assumes comparability of environments should fail. Otherwise we have to conclude that we have no evidence to support the view that environments of twins depend on zygosity.

Clearly, it would be impossible to give a full account of the specification of every possible set of circumstances for all kinds of relationships. Many of the possibilities are already in the literature (e.g. Fisher, 1918; Eaves, 1969, 1973, 1976*a*, *b*; Jinks & Fulker, 1970; Eaves & Eysenck, 1975, 1976). Here we intend to give expectations which illustrate certain basic principles of model building, for a particular set of relationships.

(*a*) *A hypothetical experiment.* We shall illustrate the principles of model building and testing by reference to an experimental design which is rather more demanding than those usually employed in psychogenetical studies in order that we may see how different effects are likely to be detected more easily in some sets of data than others. We choose the design to cover the widest range of possibilities for the causes of individual differences, whilst still restricting ourselves for convenience to those situations where the model is linear, or can easily be linearized.

We will consider the different theoretical expectations which could be applied to statistics derived from the following kinds of data:

Monozygotic twins reared together;
Dizygotic twins (full-sibs) reared together;
Monozygotic twins reared apart;
Dizygotic twins (full-siblings) reared apart;
Unrelated individuals reared together; and
Singletons, reared by their natural parents.

In practice, in order to specify correctly any model for sibling effects it is necessary to know more about the rearing conditions of those related individuals who have been reared apart. We shall give the expectations for such individuals on the assumption that they have been reared as singletons in randomly chosen foster homes. Other expectations would apply if individuals were reared with other natural or foster siblings, or reared, for example, by relatives. Similarly, our expectations on the basis of cultural transmission assume random placement. Generally speaking, placement seems to be a significant factor in many studies of IQ.

The effects of placement on the similarity between members of a family will depend on the principles on which their placement is based and on the major causes of individual differences. In most cases placement effects should produce results which are inconsistent with any simple model which assumes their absence. At one extreme, for example, if individual differences are largely inherited, placement can have little effect on the similarity of relatives reared apart, but will be reflected in a large (genetic) correlation between unrelated individuals reared together. On the other hand, if cultural effects predominate in the determination of individual differences, the differences between families of unrelated individuals reared together provide an upper limit to the effects of cultural differences and are expected to be not substantially affected by placement. When cultural effects predominate, related individuals reared apart are not expected to be more alike than unrelated individuals reared together however marked the effects of placement.

(*b*) *The basic model.* The simple model on which we build is that adopted by Jinks & Fulker (1970) in their discussion of the relationship between biometrical genetical approaches

and those proposed by Cattell in the multiple abstract variance analysis (MAVA). Though in many respects this model is not suitable in general because of its reliance on purely empirical parameters, it provides a valuable starting point for discussion, providing that we restrict ourselves to consideration of the groups of relatives we have already enumerated above. Given that all the groups of relatives represent random samples of the population of genetical and environmental influences, then in the absence of genotype–environmental interaction and genotype–environment correlation, we expect the total variances of the groups to be equal. We may then write for each group the same expectation for the total variance:

$$\sigma_p^2 = G + E.$$

This formulation is not very helpful, however, because it does not recognize the different classes of genetical and environmental variation which contribute to G and E. We may analyse G in two ways. The first approach, which is that of MAVA, is to recognize that individuals differ genetically because their particular parents bear a restricted selection of the available alleles in the population. That is, families differ genetically because their parents differ. Such differences contribute to the between-family hereditary variance, G_2 in the notation of Jinks & Fulker. Within families, however, individuals differ as a result of segregation of the alleles from the parental set. Segregation contributes to the genetical variance within families. This contribution we denote by G_1. This is one approach, which is satisfactory providing we consider only the groups of relatives here. Once we consider other kinds of relationships, parents and offspring, for example, or cousins, this approach leads to a dead end because new parameters have to be introduced to specify each new situation. Differences can always be explained in this way, but no true analysis is possible. This is the weakness of the MAVA approach. By contrast the biometrical genetical approach parameterizes the genetical components within and between families in terms of a few general parameters which specify the cumulative additive and non-additive effects of genes and which can be used to describe the contribution of genetical factors to any kind of variance or covariance in a population. When mating is random, for example, the additive and dominance effects of autosomal loci can be represented by parameters D_R and H_R respectively, thus:

$$G = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R,$$

$$G_2 = \tfrac{1}{4}D_R + \tfrac{1}{16}H_R,$$

$$G_1 = \tfrac{1}{4}D_R + \tfrac{3}{16}H_R.$$

Providing that the genetical assumptions are met, then other variances and covariances for randomly mating populations can be represented in terms of the same two parameters. We have not specified the effects of epistasis. Although this can be done (see Mather, 1974), we have not done so here because so much of the variation due to epistasis will be correlated with the factors contributing to the dominance parameter, H_R, that residual epistatic effects are likely to be too small to be detected against the background of other more significant effects; that is, although fitting only D_R will leave some epistasis in the residuals, adding H_R will account for virtually all residual non-additivity.

The effect of assortative mating is to generate linkage disequilibrium between the loci affecting a trait and to produce a slight increase in homozygosity. When many genes are involved, however, the effect of linkage disequilibrium on the genetic variance is

very much more marked than that due to increased homozygosity; indeed, the effect of the latter on individual differences is very slight. The linkage disequilibrium, however, leads to a substantial increase in genetical differences between families. Fisher (1918) showed that, if a population was in equilibrium under assortative mating, the genetical variation between families (G_2) was increased by an amount $\frac{1}{2}D_R(A/(1-A))$, where A represents the correlation between the additive genetical deviations of spouses. The relationship between A and the marital correlation will depend upon, among other things, the narrow heritability of the trait and upon the precise basis on which assortative mating takes place. The general effect of assortative mating is to make distant relatives much more alike than might be expected on the basis of a random mating model. This fact has been exploited to provide quite a powerful test of the genetical consequences of assortative mating (Eaves, 1973).

Whereas we have a ready-made theory of genetical factors which can be cast into a precise and testable mathematical form, there is a general deficiency of rigorous models for environmental variation. Although some steps have been made towards improving this situation, for example in the specification of maternal effects (Mather & Jinks, 1971), cultural effects (Cavalli-Sforza & Feldman, 1973; Eaves, 1976*b*) and sibling effects (Eaves, 1976*a*), we must be prepared to accept the fact that a large component of empiricism is likely to remain in the specification of environmental effects. When we consider the families we list above, we may distinguish two sources of environmental variation, between and within families, and we may represent their contribution to the total variance by E_2 and E_1 respectively. Whereas the genetical system imposes restraints upon the relative magnitudes of G_1 and G_2, there are no such restraints on the values of E_1 and E_2. Thus, whilst the constraint $G_1 = G_2$ is quite legitimate (since it specifies random mating and additive gene action), the constraint $E_1 = E_2$ is entirely without justification on any reasonable model for the origin of environmental differences. In general, the ratio of the two environmental components must be determined by the particular set of data in question. The fact that no *a priori* relationship exists between the environmental components can be appreciated when we ask what factors contribute to E_1 and E_2. E_1 will include such factors as developmental accidents and errors of measurement, whereas E_2 will reflect maternal and cultural effects, etc. There is no theoretical reason why the same trait should be sensitive to all kinds of environmental influence, since different kinds of influence may be critical at different stages of development. Furthermore, there is certainly no reason why the contribution of accidental factors to differences should bear any relationship at all to the contribution, for example, of maternal effects.

We have discussed the meaning of the four basic parameters which might explain our data. We now give their contributions to the second-degree statistics that could be obtained from the analysis of the six kinds of individual mentioned earlier. Given that we have paired individuals (as will be the case inevitably with twins), we may conduct a simple one-way analysis of variance and obtain mean squares within and between families (pairs). From these mean squares we could estimate the components of variance within and between families. Although for the purposes of discussion and for theoretical work it is easier to think in terms of components of variance (σ^2s), for the purposes of fitting models to data we must work with mean squares because they are independent whereas the estimated variance components are not. In the tables, therefore, we give the expectations both for the components of variance and for the mean squares based on the analysis of families of two individuals. Obviously, to obtain the expectations for mean squares based on other family sizes we simply have to substitute the corresponding expectation for the variance components multiplied by the appropriate coefficients.

In Table 2 we give the expectations of the components of variance and the mean squares for paired individuals in terms of the simple G_1, G_2, E_1, E_2 model. To obtain the expectations in terms of the components of gene action and the mating system we merely reparameterize G_1 and G_2 in terms of D_R, H_R and any additional genetical components which may be appropriate, though any model for G_1 and G_2 which involves more than two parameters will be untestable with the given set of statistics.

Table 2. Expectations of variance components and mean squares for pairs of individuals in terms of a G_1, G_2, E_1, E_2 model

	Component of variance	Mean square
Between MZT	$G_1+G_2+E_2$	$2G_1+2G_2+E_1+2E_2$
Within MZT	E_1	E_1
Between DZT	G_2+E_2	$G_1+2G_2+E_1+2E_2$
Within DZT	G_1+E_1	G_1+E_1
Between MZA	G_1+G_2	$2G_1+2G_2+E_1+E_2$
Within MZA	E_1+E_2	E_1+E_2
Between DZA	G_2	$G_1+2G_2+E_1+E_2$
Within DZA	$G_1+E_1+E_2$	$G_1+E_1+E_2$
Between UT	E_2	$G_1+G_2+E_1+2E_2$
Within UT	$G_1+G_2+E_1$	$G_1+G_2+E_1$
Singletons	$G_1+G_2+E_1+E_2$	$G_1+G_2+E_1+E_2$

5.3. *Estimating the parameters and testing the model*

Altogether our hypothetical study has generated eleven statistics. In practice, we may have fewer than these because certain groups, notably separated twins, are difficult to obtain. Various subsets of the data would enable us to estimate the parameters and still permit us to test the model. Jinks & Fulker introduced the concept of the 'minimal set of data' which would satisfy these requirements. One such set would be monozygotic twins reared together, and dizygotic twins (or full siblings) reared apart. There are many other sets, each having its own advantages and disadvantages for testing some of the assumptions implicit in the model. With eleven statistics there is a large number of alternative methods of estimating the four parameters. Indeed, the consistency of parameter estimates obtained in different ways would seem, intuitively, to be one way of testing the adequacy of the model. The method of weighted least squares, however, gives the optimal solution to the estimation problem in the sense that it gives unbiased estimates with minimum variance. When the observed mean squares are normally distributed, the estimates obtained are maximum-likelihood estimates. Provided the sample sizes are not too small, the estimates obtained by weighted least squares should be close to the maximum-likelihood estimates. The great advantage of this approach to estimation is that we not only obtain the most efficient estimates of the parameters by making use of all the available information, but we also have a test of the goodness of fit of the model which enables us to test the null hypothesis that there are no factors contributing to variation other than those specified in or confounded with factors specified in our initial model. The weighted residual sum of squares is approximately distributed as chi-square with d.f. equal to the number of mean squares less the number of parameters estimated. The method of weighted least squares is described in the genetical context in Mather & Jinks (1971), Eaves & Eysenck (1975), in Eaves (1975) for the fitting of non-linear models, and in Jinks & Fulker (1970),

although the latter authors do not iterate on the trial weight matrix. In practice, this does not make a substantial difference to the estimates provided the observations are in close agreement with their values predicted on the basis of the model, i.e. if the model fits.

6. Extending the model

If the observations obtained in a particular study do not differ significantly from those predicted on the basis of a simple model for variation there is little justification for seeking a more complex explanation by attempting to estimate additional effects not confounded with those already specified. On the other hand, if the simple model does not provide an adequate fit to the observations, we are compelled to consider a number of subsidiary theories of individual differences, including (1) genotype–environment interaction, and one or other form of (2) genotype–environment covariation. These issues are considered in turn.

6.1. *Genotype–environmental interaction*

We may incorporate additional parameters into our model to specify the effects of genotype–environmental interactions by recognizing that genetical differences within and between families may interact both with environmental differences within and between families. In principle, there are thus four possible $G \times E$ parameters which we may write:

G_1E_1 to represent the interaction of genetical and environmental influences within families;

G_2E_1 to denote the interaction of genetical differences between families with environmental differences within families;

G_1E_2 representing the interaction of genetical effects within families with environmental differences between families; and

G_2E_2 denoting the interaction of genetical and environmental differences between families.

Notice that the notation is in no way intended to imply that the interaction is in any sense a multiple of the component effects or their variances. The contributions of the four interaction components to the various components of variance and mean squares are given in Table 3. A simple 'rule of thumb' for deciding where a particular interaction term should go in the model is as follows:

> If the genetical and environmental influences contributing to an interaction both contribute separately to differences between families, i.e. to σ_B^2, then so does their interaction; otherwise, the interaction contributes to differences within families.

As a consequence of this, we will always find that G_1E_1 and G_2E_1 are confounded with E_1 and so can never be separated from it in an analysis of second-degree statistics. If such interactions have a systematic component, however, their presence may still be detected in an analysis based on third-degree statistics. Thus, the test for $G \times E$ proposed by Jinks & Fulker is a scaling test which helps us to identify causes of variation which would remain undetected in a simple analysis of variance. In contrast to this inevitable confounding of interactions involving the within-family environment, the interactions involving differences between families will be partly or wholly estimable, depending on the constellation of relatives included in the study. The interaction term G_1E_2, for example, contributes to variation between pairs of monozygotic twins reared together, but to the within-family variance for dizygotic twins reared together.

We observe too, that any interaction of the kinds we consider here will not lead to heterogeneity of the total variances for the various groups of relatives, since in every case

$$\sigma_T^2 = G_1+G_2+E_1+E_2+G_1E_1+G_2E_1+G_1E_2+G_2E_2.$$

Of course, there could be interactions of a different kind which would make the total variances unequal. If there were overall environmental differences between the groups of relatives, these could interact with genetical differences within the groups leading to heterogeneity of total variances between groups, even though the groups are representative of the population of genotypes. Under such circumstances, however, it is unlikely that such differences in variance would be unaccompanied by differences in mean between the samples. Furthermore, the pattern of heterogeneity of total variances is not expected, *a priori*, to be any of those characteristic of the basis kinds of genotype–environment covariance.

Table 3. The contribution of genotype–environmental interaction ($G \times E$) to individual differences

	Contribution to variance component		
σ^2	Genotypic	Environmental	$G \times E$
Between MZT	G_1+G_2	$+E_2$	$+G_1E_2+G_2E_2$
Within MZT		E_1	$+G_1E_1+G_2E_1$
Between DZT	G_2	$+E_2$	$+G_2E_2$
Within DZT	G_1	$+E_1$	$+G_1E_1+G_2E_1+G_1E_2$
Between MZA	G_1+G_2		
Within MZA		E_1+E_2	$+G_1E_1+G_2E_1+G_1E_2+G_2E_2$
Between DZA	G_2		
Within DZA	G_1	$+E_1+E_2$	$+G_1E_1+G_2E_1+G_1E_2+G_2E_2$
Between UT		E_2	
Within UT	G_1+G_2	$+E_1$	$+G_1E_1+G_2E_1+G_1E_2+G_2E_2$
Singletons	G_1+G_2	$+E_1+E_2$	$+G_1E_1+G_2E_1+G_1E_2+G_2E_2$

The genes which contribute to G_1 and G_2 for a particular trait need not necessarily be the same as those contributing to G_1E and G_2E interaction, and need not be associated with them genetically. Genes controlling sensitivity to the environment can often be selected quite independently of those responsible for differences in average performance over a range of environments and even located in different linkage groups (see e.g. Perkins & Jinks, 1968, 1973; Brumpton, 1973; Jinks & Connolly, 1975; Mather & Caligari, 1975). In theory, at least, we could conceive of a trait for which the model only involved terms in $G \times E$. In practice such a trait would give a readily detectable pattern of variance components providing that the study also included fostered subjects. If, however, the study was restricted to individuals reared by their natural parents, the pattern of variance components might be mistaken for that of the simple additive model. In any case, supposing we had data only on MZ and DZ twins reared together, the effects of G_2E_2 would be formally inseparable from those of G_2 and the effects of G_1E_2 would always be confounded with those of G_1. Although this is true formally, however, we should ask what it means in practice. It does not mean that the role of genetical and environmental differences is diminished. The demonstration of $G \times E$ requires only that we appreciate that genes can control sensitivity to the environment and that the environment can modulate the expression of genes. From another viewpoint, however, the demonstration of widespread $G \times E$ involving cultural

factors, especially if such interaction were of the unsystematic variety, would argue against any general procedures for the environmental modification of behaviour, since individual genotypes would respond in quite unpredictable and specific ways to changes in their cultural environment.

Our specification of $G \times E$ has so far followed purely empirical parameters since many, if not most, of the relevant considerations are model free. However, we may ask whether any reparameterization of the $G \times E$ is possible, just as we were able to represent G_1 and G_2 in terms of the components of gene action. In theory this is possible, and in practice such reparameterization may lead to a tighter and more testable model. Although we distinguish two sources of environmental variation empirically, namely E_1 and E_2, and have suggested that this distinction probably reflects real differences in the modes of environmental causation, the distinction between G_1 and G_2 is a formal one and quite arbitrary since the same genes and the same gene effects are contributing to the two genetical parameters to a degree which is simply dependent on the laws of inheritance, and can be determined mathematically. A similar relationship is to be expected for the genotype–environmental interaction. Experience with experimental organisms leads us to suppose that different genes could control sensitivity to cultural effects from those which respond to developmental accidents. That is, interactions involving E_1 and E_2 could be mediated at least in part by quite different genetical systems. (As far as the genes which determine the interaction are concerned, however, it makes no difference at all whether members of the family share the same or different alleles at a locus.) Thus, although it is legitimate to distinguish genetically and environmentally interactions involving E_1 and E_2, it makes no sense to distinguish interactions involving G_1 and G_2 except in so far as the relative contributions to additive and dominance effects differ in the two genetical components. With this in mind, therefore, we can reparameterize G_1E_2 and G_2E_2 in terms of the additive and dominance effects of genes which contribute to the interaction of genes and environmental differences between families. Thus we may write

$$G_1E_2 = \tfrac{1}{4}D_RE_2 + \tfrac{3}{16}H_RE_2 \quad \text{and} \quad G_2E_2 = \tfrac{1}{4}D_RE_2 + \tfrac{1}{16}H_RE_2,$$

where D_RE_2 represents the additive genetical component of sensitivity to the environment, and H_RE_2 denotes the variation due to dominance deviations in sensitivity to the environment. A similar reparameterization is possible for the GE_1 terms, but is of little use since all such interactions are confounded with E_1. The above expectations assume random mating with respect to the loci involved in $G \times E$. In fact, unless studies were exceptionally large and the contribution of $G \times E$ overwhelming, the chance of ever resolving $G \times E$ into its additive and dominance components must be remote in man. However, we would be content with a convincing demonstration of $G \times E$ in second-degree statistics even though an analysis of its components would be impracticable.

If there is a theory of environmental differences to match that of hereditary effects then we could offer a still more detailed specification of $G \times E$. Cavilli-Sforza & Feldman (1973) consider the case of cultural transmission acting on a single gene polymorphism and include parameters to specify 'plasticity' which is, in effect, the interaction between the family environment provided by parents and the genotype of the offspring. Such effects are components of the GE_2 interactions defined by Jinks & Fulker, and considered above in more detail.

6.2. *Genotype–environmental covariance*

We have already outlined the principle on which the analysis of Cov GE can be based. We now turn directly to its specification.

(*a*) *Individuals selecting their own environment.* Although we could produce a formal model which distinguished the direct effects of loci on an organism from those which operated indirectly by modifying the environment, such a model would have no analytical value because the components of genetical variation and genotype–environmental covariation always appear in the same expectations with the same coefficients. This is what we would expect in any situation in which the environment is merely an extension of the phenotype. This presents no more of a dilemma than the observation that fast growing genotypes eat more.

(*b*) *Sibling effects.* The specification of sibling effects has been considered in more detail, with an example, by Eaves (1976*b*). The principles and expectations are outlined here.

Just as we may represent genetical effects in terms of the additive and dominance components, D_R and H_R, so we may recognize analogous components of variance which comprise the effects of genes upon the environment, $D_{\ddot{R}}$ and $H_{\ddot{R}}$. Thus in the case of the effects of one sibling on another we may expect part of the environmental variance within and between families to be predictable from the genotypes of the individuals in the family. The loci contributing to individual differences may display additive and non-additive effects on the environment, hence the definition of a $D_{\ddot{R}}$ and an $H_{\ddot{R}}$. In general, there is no particular reason why the genes which contribute to the 'ordinary' genetical variance (represented by D_R and H_R) should bear any functional or spatial relationship (in effect or location) to those which affect the environment (and contribute to that part of the environmental variation represented by $D_{\ddot{R}}$ and $H_{\ddot{R}}$). When the two kinds of gene effect are independent D_R and $D_{\ddot{R}}$ will remain confounded as long as the individuals studied are all reared at the same density (e.g. in pairs). In the absence of $G \times E$, we might expect singletons, for example, to show a smaller variance on account of their not being exposed to this additional source of environmental variation.

When, however, the genes which affect the environment are associated (e.g. as a result of pleiotropy or linkage disequilibrium) with those which affect the phenotype of individuals directly, we have the possibility of genotype–environmental covariation. The net contribution of such common effects can be represented by two genotype–environmental covariance parameters, $D_{\dot{R}}$ and $H_{\dot{R}}$. In situations where an advantaged genotype also promotes the performance of his sibling, we may speak of 'cooperation' and would expect $D_{\dot{R}}$ and $H_{\dot{R}}$ to be positive. When an individual succeeds to the detriment of his sibling we may speak of 'competition' and would expect the two covariance parameters to take significant negative values. Although, for families raised at the same density, D_R and $D_{\ddot{R}}$ are confounded, as are $H_{\ddot{R}}$ and H_R, this is not the case for the covariance parameters since their relative contributions to the total variances and to the components of variance within and between families will depend on the degree of relationship between the competing or cooperating individuals. The effect of $D_{\dot{R}}$ and $H_{\dot{R}}$ on the total variance will depend on the intensity of competition or cooperation, and this in turn will increase with increasing genotypic similarity of the individuals. Thus, for identical twins, we would expect the variance to be greater than that for non-identical twins in the presence of cooperation and less in the case of competition.

We give the expectations for the eleven statistics we are considering in Table 4, and remark that competition and cooperation give rise to a very striking pattern of individual differences which should be unmistakable in practice. The effect of competition is likely to be particularly obvious since intense competition could give rise to negative covariance between individuals in the same family, especially when the competing individuals are unrelated. We suspect that many of the extant studies

of behaviour in twins would be illuminated if the raw data could be re-examined in the light of the tabulated model. The details of Table 4 add weight to the earlier comment that the traditional specifications of Cov GE lead nowhere. We see that the part of the environmental variation which is a direct consequence of sibling effects is inevitably confounded with the genetical variation, whilst any environmental variation which could be separated from the genetical variation is functionally irrelevant to the genotype–environment covariance. Because the MAVA definitions do not make explicit the different sources of environmental variance, any attempt to specify genotype–environmental correlations through the MAVA approach is destined to give a series of seemingly unrelated and inappropriate coefficients which have no analytical or predictive value.

Table 4. The contribution of sibling effects to individual differences

σ^2	Genetical variance	Environmental variance	Genotype–environmental covariance
Between MZT	$\frac{1}{2}D_R+\frac{1}{4}H_R$	$\frac{1}{2}D_{\ddot{R}}+\frac{1}{4}H_{\ddot{R}}$	$D_{\dot{R}}+\frac{1}{2}H_{\dot{R}}$
Within MZT	—	—	—
Between DZT	$\frac{1}{4}D_R+\frac{1}{16}H_R$	$\frac{1}{4}D_{\ddot{R}}+\frac{1}{16}H_{\ddot{R}}$	$D_{\dot{R}}+\frac{1}{2}H_{\dot{R}}$
Within DZT	$\frac{1}{4}D_R+\frac{3}{16}H_R$	$\frac{1}{4}D_{\ddot{R}}+\frac{3}{16}H_{\ddot{R}}$	$-\frac{1}{2}D_{\dot{R}}-\frac{1}{8}H_{\dot{R}}$
Between MZA	$\frac{1}{2}D_R+\frac{1}{4}H_R$	—	—
Within MZA	—	—	—
Between DZA	$\frac{1}{4}D_R+\frac{1}{16}H_R$	—	—
Within DZA	$\frac{1}{4}D_R+\frac{3}{16}H_R$	—	—
Between UT	—	—	$D_{\dot{R}}+\frac{1}{2}H_{\dot{R}}$
Within UT	$\frac{1}{2}D_R+\frac{1}{4}H_R$	$\frac{1}{2}D_{\ddot{R}}+\frac{1}{4}H_{\ddot{R}}$	$-D_{\dot{R}}-\frac{1}{2}H_{\dot{R}}$
Singletons	$\frac{1}{2}D_R+\frac{1}{4}H_R$	—	—

(*c*) *Cultural transmission.* If the environment shared by offspring depends on the phenotype of their parents, then we expect genotype–environmental covariance whenever the genes which influence the offspring directly are correlated in their effects or distribution with those which determine the quality of environment provided by the parents. This effect, which may be called 'cultural transmission', produces covariance between genetical and environmental differences between families (Cov $g_2 e_2$) which will appear in the expectations of any between-family component of variance obtained for individuals reared by their natural parents. Whenever individuals are reared in foster homes this covariance is reduced to zero (if fostering is random), with the result that the variance of fostered individuals is expected to be less than that of individuals reared by their natural parents, if the covariance is positive.
Jinks & Fulker (1970) proposed comparing the total variances of twins and siblings reared together with that of twins and siblings reared apart as a simple scaling test of Cov $g_2 e_2$. Two bodies of data illustrate the principle quite well. The 53 pairs of separated MZ twins studied by Burt (1966) can be divided into those which were reared in their own home and those which were reared in foster homes. There is no mean difference in IQ between the two groups and no covariance between the SES of the natural and foster homes. Unfortunately, we have only 53 pairs, so the test is not very powerful, but the variance of individuals raised in their own home is 215·81, whilst the

variance of the twins raised in foster homes in 231·44. These two variances do not differ significantly, which suggests that we have no reason to suppose Cov $g_2 e_2$ is a significant factor in individual differences in these data. Similarly, we have data on the EPQ psychoticism scores of individuals reared by their natural parents or by foster parents. The variance of 1153 individuals reared by their natural parents was 0·02766 whilst that of 340 individuals reared by foster parents was 0·02570. Here the sample sizes are large enough to provide a fairly powerful test of a major genotype–environmental covariance component. Since the variances do not differ significantly we must conclude that Cov $g_2 e_2$ can make only a relatively small contribution to individual differences in psychoticism, as it is measured by the P scale of the EPQ. This confirms a tentative conclusion drawn from twin data by Eaves & Eysenck (1977) that parents make no contribution to the development of psychoticism in their children above that provided by their genes. Eaves (1976*b*) gives expectations for various collateral and ancestral relationships in terms of a model for cultural transmission which involves polygenic inheritance, random mating and culturally transmissible environmental 'accidents'. Under such circumstances the environmental component of differences between families can be represented in terms of the sources of genetical and environmental differences which would persist if cultural transmission were terminated. Similarly, the genotype–environmental covariance can be expressed for a given intensity of cultural transmission, in terms of the additive effects of genes.

Table 5. The general effect of cultural transmission on individual differences

σ^2	Genetical variance	Environmental variance	Genotype–environmental covariance
Between MZT	G_1+G_2	$+E_2$	$+2$ Cov $g_2 e_2$
Within MZT		E_1	
Between DZT	G_2	$+E_2$	$+2$ Cov $g_2 e_2$
Within DZT	G_1	$+E_1$	
Between MZA	G_1+G_2		
Within MZA		E_1+E_2	
Between DZA	G_2		
Within DZA	G_1	$+E_1+E_2$	
Between UT		E_2	
Within UT	G_1+G_2	$+E_1$	
Singletons	G_1+G_2	$+E_1+E_2$	$+2$ Cov $g_2 e_2$

In Table 5, the contributions of cultural effects of Cov GE are given in terms of the empirical parameters Cov $g_2 e_2$ rather than in terms of any particular theory of cultural transmission, since we wish to illustrate the point that the scaling test of genotype–environment covariance between families provided by a comparison of total variances does not depend on any particular set of assumptions about the mating system or the mode of gene action. The parameterization of E_2 and Cov $g_2 e_2$ in terms of more explicit models for cultural transmission is discussed by Cavalli-Sforza & Feldman (1973) and Eaves (1976*b*). Here, random placement has been assumed for fostered individuals and all other individuals are assumed to have been reared by their natural parents. When we turn, later, to a study of the power of the test, we shall provide an illustration of how this parameter may be represented in terms of a stronger theory of individual differences.

Some authors (e.g. Moran, 1973) have taken pains to assert that the presence of genotype–environment covariance would appear to undermine attempts to estimate the contribution of genetical factors to individual differences. Clearly, in the presence of sibling effects and cultural transmission, our model for individual differences will be more complex than that which is the basis of the simplest estimates of broad and narrow heritability. What such authors with their highly conservative approach fail to indicate, however, is the great biological, indeed genetical and psychological significance of $\mathrm{Cov}\,GE$, and hence the importance of detecting it and understanding it. If we detect $\mathrm{Cov}\,GE$ we are, in effect, claiming that part, at least, of what was once termed 'environmental' variance is a reflection of genetical polymorphism and may be influenced by natural selection. Just as, in the presence of $G \times E$, we can ask quite legitimately 'To what extent are differences in sensitivity to the environment under genetical control?', so in the presence of $\mathrm{Cov}\,GE$ we are forced to ask 'To what extent are differences in the environment under genetical control?'. Just as, in the presence of $G \times E$, our estimate of the significance of genetical factors depends on the environment in which the measurements are made, so it is also the case with $\mathrm{Cov}\,GE$. In the presence of sibling effects, for example, we cannot obtain a single estimate of the significance of genetical factors for the population as a whole, but we can still estimate the significance of genetical factors for each of the different rearing systems. Eaves & Eysenck (1977) have illustrated the relative strength of this approach in their analysis of data relating to psychoticism. A further illustration of the method is given by Martin & Eysenck (1976) for sexual satisfaction in females.

7. Scaling tests for the detection of systematic non-additive effects

We have dealt with the specification of second-degree statistics first because our primary aim is to understand variation and to provide a theoretical framework for more rigorous discourse in this area. We have, however, already made the distinction, of some theoretical and practical significance, between the systematic and unsystematic (directional and ambi-directional) effects, and we have suggested that even if the approach of model fitting to second-degree statistics could detect non-additivity, it would be incapable of resolving directional and ambi-directional effects. We will now, therefore, consider the various methods of analysis which might help us detect systematic genetical and genotype–environmental interactions.

7.1. *Detecting genotype–environment interaction*

Jinks & Fulker (1970) suggested that certain kinds of $G \times E$ could be detected by inspecting the form of any relationship between the means and absolute intrapair differences (or within-pair variances) for MZ twins. The principle behind this approach has been often misunderstood but stems from the recognition that, in experimental organisms, a major component of $G \times E$ is due to the effects of genes which are functionally or spatially associated with those responsible for overall differences in a trait. Whilst such 'systematic' effects do not exhaust all the possibilities for $G \times E$ they turn out, in practice, to be a major component of $G \times E$ and their proper analysis is crucial to any prediction of performance. In the presence of such interactions a change of scale might be indicated to secure additivity but often this is impossible because the original scale has other desirable properties which would be lost on transformation.

Consider an array of n genotypes grown in two contrasting environments.

A conventional analysis of variance would recognize, in addition to replicate error, the following items:

Item	d.f.
Between genotypes	$n-1$
Between environments	1
Genotypes x environments	$n-1$

Such an analysis, however, is only the starting point since the $G \times E$ term contains a multitude of effects, some of which may be identified and employed to enhance prediction. Biometrical geneticists (e.g. Perkins & Jinks, 1968, 1973; Bucio-Alanis *et al.*, 1969) have shown how a systematic component of $G \times E$ can be isolated, if such is present, by recognizing that the heterogeneity of the n differences over environments may regress significantly on the overall expression of a genotype in the chosen environments. Thus, we may divide the $(n-1)$ d.f. for $G \times E$ into two further components:

Item	d.f.
Regression on genotypic mean	1
Residual (unsystematic) $G \times E$	$n-2$

In practice, it is often found that selection can alter the pattern of association between genotypic mean and deviations due to $G \times E$. That is, relationships which might be removed by rescaling the data are themselves a reflection of the association (presumably as a result of selection) between genes which affect overall expression of a trait, and those which control responsiveness to environmental differences. The test proposed by Jinks & Fulker, and a variety of related tests, share much in common with this generally approved and successful approach to the analysis of $G \times E$. Consider, for example, as Jinks & Fulker did, the analysis of n pairs of identical twins reared apart (MZA). A conventional analysis of variance of such data would recognize the following items:

Between pairs $n-1$

Within pairs n

The sum of squares within pairs will incorporate differences due to errors of measurement, differences due to genuine environmental effects and variation due to genotype–environmental interaction. As long as environmental differences are random, and $G \times E$ is unsystematic, there can be no relationship between the differences within pairs and the pair means. This will apply whatever the genetical system contributing to differences between pairs.

In every pair of twins, one twin will score higher than the other for any trait that is subject to some degree of environmental control. In so far as this difference is due to chance there can be no relationship between the magnitude of the intrapair difference and the overall deviation of the pair from the population mean (given uniform error

variance throughout the scale). The difference in score, however, implies that relative to one another, one twin has experienced a superior environment and the other an inferior environment. We may use the difference in scores within a pair as an indicator (of uncertain realiability) of the difference between superior and inferior environments, and we can ask: 'Does the difference between the effect of superior and inferior environments depend on the genotype of the individuals concerned?'. Jinks & Fulker approach the question through a regression analysis of the absolute intrapair differences. The uncorrected sum of squares of the absolute differences is equal to the within-pair sum of squares of the conventional analysis of variance of the paired scores. Implicit in the regression analysis is the following partition of this sum of squares:

Item	d.f.
Correction term ('superior vs. inferior environment')	1
Linear regression (systematic $G \times E$) on pair means	1
Residual (non-linear systematic $G \times E$, unsystematic $G \times E$, etc.)	$n-2$

The linear regression on pair means will be significant if there is $G \times E$ such that the genes which contribute to sensitivity to the environment are correlated in distribution or action with those affecting overall performance. We may conceive of such interactions in physiological or social terms. Thus, for example, individuals who score high on a trait may be more stable developmentally than low scorers. That is, the effect of the same environmental influences changes systematically with genotype. Alternatively, individuals who on average manifest high scores may find themselves in more variable environments because there is no consistent policy towards the treatment of such individuals (Eaves, 1970).

We may illustrate the approach by reference once more to the often publicized and much criticized data of Burt (1966), described in section 6.2(c) above, which comprises IQ scores of 53 pairs of MZ twins. One member of each pair was reared by his natural parents and the other in a foster home. There is no detectable correlation between indices of SES of the natural and foster homes.

The basic analysis of variance of these data is as follows:

Item	d.f.	SS	MS
Between pairs	52	21 817·60	419·57
Within pairs	53	1 455·00	27·45

Following the argument of Jinks & Fulker (1970) we may further analyse the variation within pairs as follows:

Item	d.f.	SS	MS	% of total SS within pairs
'Superior vs. inferior environment'	1	942·04	942·04	64·7
Regression on mean	1	2·09	2·09	0·1
Residual	51	510·87	10·02	35·1
Total	53	1455·00	—	99·9

The comparison of 'superior and inferior' environments, based as it is upon an examination of the twins' scores, is likely to be an overestimate of the 'true' overall difference between the twins' environments and is included merely to indicate the structure of the analysis. The residual sum of squares is a guide to the relative importance of heterogeneity of superior and inferior environments over pairs and unsystematic $G \times E$ effects.

If we assume, for the moment, that the usual tests of significance are appropriate here, we find that the regression does not account for a significant proportion of the variation in environmental differences around their overall mean. That is, the data give us no reason to suppose that systematic $G \times E$ effects need to be taken into account when making predictions on the basis of IQ tests.

We may wonder whether the test of significance is appropriate given that we decide from the data which twin received which environment. Dr M. J. Kearsey of our Department kindly made available to us a computer package he had developed for the simulation of various aspects of polygenic inheritance and experimental design in biometrical genetics. Genetical variation is simulated on the assumption of ten loci in linkage equilibrium. The dominance effects of the loci and the frequencies of their alleles can be varied. Random mating is simulated. Environmental variation is simulated by random processes, either specific to individuals or common to families. As part of a programme of computer simulation (Martin, 1976), we generated 400 sets of 500 MZ twin pairs. This comprised 25 replications of each of eight different populations, differing in the degree of genetic determination, the amount of dominance and the relative importance of environmental differences between families. No systematic $G \times E$ was specified in the model used to generate the data. The regression analysis described above was used for the 400 sets of twins and in only 14 of the simulations ($3\frac{1}{2}$ per cent) was a significant linear component of regression detected at the 5 per cent level. The fact that this is close to the expected proportion of significant results, given that the null hypothesis is true, suggests that the test of significance we employ is not seriously misleading. These simulations also confirm the model of Jinks & Fulker which implies that in the absence of $G \times E$ there is expected to be no significant relationship between sum and difference for monozygotic twins reared apart.

The power of the test for systematic G × E. Although, as we shall show, small amounts of unsystematic $G \times E$ are difficult to detect, our experience has shown that this is not the case for systematic types of $G \times E$. In an analysis of a variety of behavioural traits involving quite small samples of MZ twins reared together, Martin (1976) showed that systematic non-additive effects, accounting for no more than 5 per cent of the total variation within MZ pairs, could be detected at the 5 per cent level with only 95 pairs of twins. Even with samples as small as 39 pairs he was able to detect non-additive effects accounting for not more than 12 per cent of the variation within pairs.

So far we have restricted our consideration of $G \times E$ to systematic linear effects. There is no particular reason why such effects should be linearly related to the genotypic mean. Indeed, there are many situations in which we might find significant non-linear trends. Society may react in a uniform way to extreme deviations on either side of the population mean. This would produce a pattern of $G \times E$ which shows greater environmental variation in the middle of the scale than at either end. In practice, this kind of interaction is common in psychometric data because of floor and ceiling effects. In Burt's data there is little suggestion of such interaction, but many personality scales show apparent interactions of this type when raw scores are used. In a recent twin study, responses of 316 MZ twins were obtained to 74 items relating to neuroticism.

When neuroticism scores were computed simply by counting the number of 'neurotic' responses made by each individual, a highly significant quadratic relationship was obtained, with pairs at the tails showing much smaller differences within pairs than those in the middle of the range.

Genotype–environmental interaction and 'scale'. Such interaction could be explained in terms of the properties of the scale, since it is well known that the error variance of scores of this type depends on the subject's propensity to endorse dichotomous items in a given direction. In order to test the hypothesis that all the $G \times E$ interaction could be attributed to the heteroscedasticity created by a choice of scale, we attempted to remove the interaction by an angular transformation of the proportional scores. Indeed, we found that the transformed scale showed no evidence of any systematic non-additivity of any kind. In a similar study of psychoticism, Eaves & Eysenck (1977) showed that any non-additivity could be removed by a simple transformation which assumed that the psychotic responses of individuals to items of the EPQ were randomly distributed over the items of the P scale with a frequency for each subject which appeared to be under genetical control. Last (1976) has analysed raw scores relating to various aspects of cognitive abilities for black and white twins in the US. She found similar interactions to those found for personality scores, and concluded that sexes and races might differ superficially in their apparent sensitivity to differences in the environment, but that such apparent differences can largely be attributed to the fact that threshold effects play a more significant role in some groups than others.

These findings point to a dilemma in the analysis of certain behavioural data. Either we accept the scale that is used for convenience and tolerate the non-additivity generated by such scales, or we attempt to produce a scale which is free of such interaction but which will inevitably be measuring a somewhat different trait. These scales permit a more reliable analysis, produce tidier conclusions and have much to commend them for predictive purposes because the problem of heteroscedasticity has been overcome. The search for a scale which is more appealing from the psychological standpoint is not, however, without its problems from the biometrical–genetical view.

We may illustrate the problem by brief reference to an attempt we made to measure neuroticism on a scale which had greater behavioural relevance. We took the responses of 1174 twins of both sexes (including our 316 MZ twin pairs) to the 74 neuroticism items of our experimental questionnaire and subjected these to analysis in terms of the two-parameter logistic model of Birnbaum described by him in Lord & Novick (1968). We estimated the item parameters and the latent trait scores for all the subjects in the study. In spite of the fact that we had a fairly large number of items related to neuroticism, at least by comparison with those in the N scales of the EPQ and EPI, the amount of information (i.e. the inverse of the score variance) computed for different scores on the latent trait model showed a very substantial reduction as scores became more extreme. In fact, information was maximal (or the variance of scores was smallest) for scores approximately 1 S.D. from the zero point of the scale. This is to be expected for a collection of items which have presumably been selected at some time or other for the discrimination of relatively extreme neurotics from the rest of the population.

Whilst such a model makes psychological sense, its consequences for any genetical analysis are quite extreme. The relationship between sums and differences for the latent trait scores shows the marked U-shaped relationship to be expected from the fact that the amount of information about a subject is so greatly reduced towards the extremes of the distribution. Of all the possible ways of summarizing the information in a set of

responses to the same questionnaire data, that which is most appealing for psychometric reasons is that which requires the most complicated explanation of the pattern of individual differences. Indeed, none of the simpler additive models could account for the variation in neuroticism as it was measured by a logistic model for the subjects' responses. In fact, the genotype–environmental interaction imparted to the data an apparent consistency with a mechanism involving some kind of sex linkage or sex limitation, to judge from the differences in mean and variance between the sexes. A similar result has been claimed by others (e.g. Bock & Kolakowski, 1973) for variation in spatial ability when subjects' scores have been assessed on the basis of a similar psychometric approach. A recent study of parent offspring similarity by DeFries *et al.* (1976) does not support the hypothesis of sex linkage for spatial visualization and leads us to wonder whether scalar properties could not explain the earlier finding.

The important point to notice about this discussion is that three different methods of scaling the same behavioural data yield quite different conclusions about the form of $G \times E$ (and presumably also about the genetic architecture). The raw scores yield a scale on which individuals at the extremes are apparently less sensitive to the environment, the angular transformation yields a scale on which environmental factors have a uniform effect over the whole range, and the logistic model yields a scale on which the influence of the environment appears to be much greater in the tails. None of these scales is 'right', they all employ the same information. As far as a genetical analysis is concerned a change of scale amounts to stressing the consequences of some gene substitutions or environmental influences at the expense of others. In the end, a satisfactory scale is one which yields the most effective general predictions.

There is no reason to suppose that all scales show such striking problems of non-additivity. In an analysis of a social attitude questionnaire consisting of 68 trichotomous items scored 1, 2, 3 we found that a general factor corresponding to 'conservatism' produced scores for the 316 MZ twin pairs which showed virtually no evidence of non-additive effects of the kind we have discussed so far. Out of the 14 ability scales studied by Last (1976), six showed evidence of systematic $G \times E$ in their raw form. Often simple transformations were insufficient to remove such non-additive effects. Martin (1976) has analysed the raw scores of 134 MZ twins pairs from a variety of personality and attitude factors and concluded that five out of ten scales showed some form of systematic non-additivity which could be removed by suitable transformation. In a reanalysis of certain data on 40 MZ pairs from the Michigan twin study (Vandenberg, 1962), Eaves (1970) found systematic *linear* interactions for 12 out of 36 traits covering a wide range of behavioural measurement.

Jinks & Fulker (1970) considered the consequences of $G \times E$ for the relationships between pair means and absolute differences for MZ twins. They observe that twins reared apart will provide the best basis for detection of $G \times E$ for two reasons:

(1) All $G \times E$ involving interaction of genetical differences with environmental differences within and between families will be confounded with differences within pairs, whereas only genetical differences will contribute to differences between pairs, given random placement. Such a set of twins, therefore, means we can study the components of *all* kinds of $G \times E$ in relation to genetical differences in the population. If the twins are reared together, however, any interaction of genetical differences with between-family environmental effects is confounded with the corresponding genetical and environmental effects. This means that MZ twins reared together can only be used to detect systematic components of interaction between genes and intrafamily environmental effects.

(2) The use of MZA will ensure, as far as possible, that interaction between different sources of environmental variation will contribute only to differences within pairs and be

uncorrelated with genetical effects. Interaction of this type would result if there were cultural differences between families such that certain parents tended to treat all their own children uniformly whilst other parents treated their children individually.

Whatever the family structure, interactions which involve differences within families will contribute to variation within families, but if there is systematic interaction between cultural effects and treatment differences within families ($E_2 \times E_1$) these will contribute, as Jinks & Fulker have shown, to the covariance of means and intrapair absolute differences for MZT. In order to interpret a relationship between means and differences for MZT as $G \times E$ we have to be fairly sure that there is no major contribution from cultural effects to differences between pairs. Ideally such a question is best answered by data on fostered individuals, although the approach of model fitting to statistics on relatives reared together can give some indication of the likely significance of cultural differences between families (see e.g. Eaves & Eysenck, 1977).

7.2. *Detecting systematic genetical non-additivity*

Our test of $G \times E$ is effectively a test based on third-degree statistics since we are looking at the covariance between means and a measure of variation, in this case the standard error, within families. In general, we would expect the effects of other systematic non-additive effects to be detectable with third-degree statistics. In any study of individual differences there is a variety of such statistics that can be calculated which may shed additional light on sources of non-additive variation. Some, but not all, of these possibilities were considered by Jinks & Fulker (1970). For example, they noted that the average skewness within sibling families could detect directional dominance in the absence of $G \times E$. Fisher *et al.* (1932) showed, for the case of equal allele frequencies, that a variety of third-degree statistics could provide information about the direction of dominance in the absence of other sources of systematic non-additivity. Thus the population skewness was expected to differ from zero in the presence of directional dominance and the means and variances of sibships were expected to covary. As the number of loci increases, the contribution of dominance to third-degree statistics is reduced. In natural populations, for which allele frequencies are not expected to be equal, any systematic inequality of allele frequencies (i.e. any general tendency for increasing alleles to be more or less frequent than decreasing alleles) will contribute to the same third-degree statistics, even when gene action is additive. Similarly, any general systematic epistatic component, such as that introduced because of a threshold effect on gene substitution, will result in detectable skewness or covariance of family means and variances.

The power of the test. Martin (1976) used the simulation package developed by Kearsey and described in section 7.1 to test the practical utility of these approaches to the detection of systematic genetical effects. From a variety of such populations Martin generated 25 replicate sets of 500 pairs of siblings and tested for population skewness and for the regression of absolute intrapair differences on pair means. A sample size of 500 pairs was chosen as one which might reasonably be expected in practice.

Some of the salient conclusions for this study are given in Table 6. This table includes only those simulations in which the allele frequencies were 0·5 at every locus. In such cases, any skewness or covariance of pair means and variances is attributable to directional dominance. We see that the power of the test of directional dominance, for the given sample size, is quite good (80 per cent with 500 pairs) provided that the broad heritability is high. As we might expect, however, the power of the test falls

quite rapidly as the contribution of environmental variation increases. Eaves (1972) obtained similar results relating to the detection of dominance by model fitting to second-degree statistics. In that case, however, the emphasis was on the detection of dominance variation, irrespective of any directional component. As we might expect, a weaker null hypothesis will be accompanied by a reduction of our ability to reject such a null hypothesis given that it is actually false. Thus, we find that if we do not specify the direction of dominance, but are merely concerned to detect variation due to dominance, considerably larger sample sizes are needed.

Table 6. The power of detection (5 per cent level) of directional dominance with third-degree statistics from simulated twin studies using 25 replications†

h_b^2	h/d	Cultural effects	Expected skewness	Observed skewness	Observed power (%)
0·9	1·0	Absent	−0·31	−0·31	96
		Present	−0·31	−0·30	100
	0·5	Absent	−0·24	−0·24	84
		Present	−0·24	−0·22	80
0·5	1·0	Absent	−0·13	−0·14	48
		Present	−0·13	−0·13	32
	0·5	Absent	−0·10	−0·09	12
		Present	−0·10	−0·09	16

† h_b^2 is the broad heritability; h/d is the ratio of dominance deviations to additive effects.

Martin (1976) simulated data on monozygotic twins in addition to those on siblings for the same set of simulated populations. After summarizing the data by analysis of variance within and between pairs, estimates of additive, dominance and environmental variation were obtained with their standard errors. The power of the test of dominance variation, based on 25 replicates of 500 MZ and 500 sibling pairs, was found empirically to be 80 per cent for the population with complete dominance and a broad heritability of 0·9, but when the dominance ratio was reduced to 0·5 the power was only 28 per cent. For populations in which the broad heritability was 0·5 the power of the test of dominance was virtually zero with these sample sizes. These findings confirm those obtained by Eaves (1972) for the expected power of tests of dominance by model fitting to second-degree statistics.

8. Detecting unsystematic $G \times E$ and genotype–environmental covariation

The theoretical investigation of Last (1976) is too extensive to consider in all its details, but we abstract the features of her approach which are most salient in our context. Starting with the experimental design we have introduced, she produced expected values for the various second-degree statistics obtained from such a study for a variety of systems of causation. Then, using the approach of weighted least squares to obtain the minimum variance unbiased estimates, she determined the total sample size necessary to be 95 per cent certain of detecting given effects at the 5 per cent level, given that the correct model was fitted. The general method she used was that of Eaves (1972). However, she went further to investigate the biases introduced into estimates of the parameters of various inappropriate models, given that such models were mistakenly adopted as provisional explanations for the observed pattern of

individual differences. Also, by looking at the significance of the residuals after fitting inappropriate models she was further able to judge the likelihood of detecting the mistake. The models employed covered many situations involving various degrees of non additivity and non-independence of genetical and environmental effects.

The experimental design of section 5.2(a) was adopted to allow all the effects in which we were interested to be detected, in principle, within the framework of linear model fitting. Many effects, for example, those of assortative mating and cultural transmission, make contributions to other second-degree statistics which are formulated more parsimoniously in terms of a non-linear model. The application of weighted least squares to the testing of such models for real data is quite possible (see e.g. Eaves, 1975), but the approach is tedious for simulation studies such as these.

8.1. *Genotype–environment interaction*

To help the reader grasp the method we give one example in more detail before outlining the results of others. Consider a randomly mating population in which there is additive gene action and environmental variation within and between families. The environmental differences between families are independent of genetical differences but the two sources of variation interact producing variation due to the interaction of additive genetical differences with environmental differences between families. The model for the population variance may then be written as

$$\sigma_p^2 = \tfrac{1}{2}D_R + E_1 + E_2 + \tfrac{1}{2}(D_R E_2).$$

As previously, the term $D_R E_2$ does not denote a multiple, but indicates the interaction of additive genetical differences with environmental differences between families.

In general, the genes and environments which contribute to $D_R E_2$ need not also contribute to D_R and E_2, but if they do contribute in any systematic way to these parameters, then their effects could be detected with some of the methods considered above.

(*a*) *Generating the 'data'*. Expectations for the 11 statistics of our experimental study were written in terms of these parameters. Parameters defining interactions involving within-family environmental differences were not included since these are inevitably confounded with E_1. The expectations are given in Table 7.

In order to generate numerical expectations for the 11 statistics we must specify values for the four parameters. These are arbitrary, but some sets of values are more consistent with the existing body of information about the causes of continuous variation. Thus, it seems reasonable to suppose that about half of the total variation for a trait is due to environmental differences. So for convenience, fixing σ_p^2 at 225 we have

$$E_1 + E_2 = 225/2 = 112{\cdot}5.$$

The contributions of E_1 and E_2 are determined empirically for any real set of data. We shall generate expectations on the assumption that $E_1 = E_2 = 56{\cdot}25$. The remaining 50 per cent of the variance can be assigned to the additive effects of genes and their interaction with environmental differences between families; that is, $\frac{1}{2}D_R + \frac{1}{2}D_R E_2 = 112{\cdot}5$. We now must decide the relative magnitudes of D_R and $D_R E_2$ to be incorporated in our simulations. In experimental organisms it is seldom found that more than 20 per cent of the variation between genotypes and treatments can be attributed to $G \times E$, so we choose a value of $D_R E_2$ satisfying $\frac{1}{2}D_R E_2 = 225/5 = 45$. Thus, we give the value 90

to $D_R E_2$ and to D_R we give the value $(112{\cdot}5 - 45) \times 2 = 135$. The vector of parameters used to generate the expected statistics, for the purposes of the present example, is thus:

$$\begin{pmatrix} D_R \\ D_R E_2 \\ E_2 \\ E_1 \end{pmatrix} = \begin{pmatrix} 135{\cdot}00 \\ 90{\cdot}00 \\ 56{\cdot}25 \\ 56{\cdot}25 \end{pmatrix}.$$

The expected statistics are obtained by post-multiplying the model matrix of Table 7 by this vector of estimates. For example, the expected mean square between MZ pairs reared together is

$$\begin{aligned} \mathrm{MS}_{b\mathrm{MZT}} &= D_R + D_R E_2 + E_1 + 2E_2 \\ &= 135{\cdot}00 + 90{\cdot}00 + 56{\cdot}25 + 112{\cdot}50 \\ &= 393{\cdot}75. \end{aligned}$$

The full vector of expected mean squares, given our parameter values, is given in the same table.

Table 7. A simple model involving genotype–environmental interactions used for simulation

Mean square	D_R	E_1	E_2	$D_R E_2$
Between MZT	1	1	2	1
Within MZT	·	1	·	·
Between DZT	$\frac{3}{4}$	1	2	$\frac{3}{4}$
Within DZT	$\frac{1}{4}$	1	·	$\frac{1}{4}$
Between MZA	1	1	1	$\frac{1}{2}$
Within MZA	·	1	1	$\frac{1}{2}$
Between DZA	$\frac{3}{4}$	1	1	$\frac{1}{2}$
Within DZA	$\frac{1}{4}$	1	1	$\frac{1}{2}$
Between UT	$\frac{1}{2}$	1	2	$\frac{1}{2}$
Within UT	$\frac{1}{2}$	1	·	$\frac{1}{2}$
Singletons	$\frac{1}{2}$	1	1	$\frac{1}{2}$

(*b*) *The 'estimates' and their expected variances.* Now we can do several useful computations. For example, we may proceed as if we were fitting the correct model (i.e. the model in Table 7) and obtain expected values for the variances and covariances of the four parameters. The covariance matrix of the estimates would be

$$\mathbf{Z} = (\mathbf{A}'\mathbf{W}\mathbf{A})^{-1},$$

where **A** is the model matrix (Table 7) and the diagonal elements of **W** are the reciprocals of the expected variances of the observed mean squares, i.e.

$$w_{ii} = n_i / 2\varepsilon x_i^2$$

where the n_i are the d.f. of the mean squares x_i. We have made one approximation for simplicity, by assuming all the n_i are equal. In fact, the d.f. between pairs should be one less than those within pairs. The simplification makes the repetition of the calculations easier, and has a trivial effect on the outcome.

The diagonal elements of **W** are: (0·000322, 0·015802, 0·000584, 0·002016, 0·000439, 0·003951, 0·000747, 0·001367, 0·000632, 0·001756, 0·000988). Since the mean squares are independent, all off-diagonal elements are zero.

The covariance matrix is

$$\begin{array}{c} \\ D_R \\ D_R E_2 \\ E_2 \\ E_1 \end{array} \begin{array}{c} \begin{array}{cccc} D_R & D_R E_2 & E_2 & E_1 \end{array} \\ \begin{pmatrix} 1409{\cdot}6177 & -1304{\cdot}6148 & 61{\cdot}3961 & 3{\cdot}8203 \\ -1304{\cdot}6148 & 2841{\cdot}1789 & -635{\cdot}0964 & -156{\cdot}5776 \\ 61{\cdot}3961 & -635{\cdot}0964 & 321{\cdot}5570 & 20{\cdot}4463 \\ 3{\cdot}8203 & -156{\cdot}5776 & 20{\cdot}4463 & 59{\cdot}3642 \end{pmatrix} \end{array}.$$

The estimates ($\boldsymbol{\theta}$) and their standard errors ($\boldsymbol{\sigma}_\theta$) are thus:

$$\begin{array}{c} \begin{pmatrix} D_R \\ D_R E_2 \\ E_2 \\ E_1 \end{pmatrix} \\ \\ \end{array} \begin{array}{c} \begin{pmatrix} 135{\cdot}00 \\ 90{\cdot}00 \\ 56{\cdot}25 \\ 56{\cdot}25 \end{pmatrix} \\ \boldsymbol{\theta} \end{array} \begin{array}{c} \begin{pmatrix} 37{\cdot}54 \\ 53{\cdot}30 \\ 17{\cdot}93 \\ 7{\cdot}70 \end{pmatrix} \\ \boldsymbol{\sigma}_\theta \end{array}.$$

We may divide each element of $\boldsymbol{\theta}$ by its corresponding σ_θ to give a vector of normal deviates which are the expected values of $\theta_i/\sigma_{\theta i}$ for samples of this size and structure from populations in which individual differences are determined by the effects specified in the model. This vector is

$$\mathbf{c} = \begin{pmatrix} 3{\cdot}60 \\ 1{\cdot}69 \\ 3{\cdot}14 \\ 7{\cdot}30 \end{pmatrix}$$

In practice we would perform a two-tailed test of significance because, even though the components of variance are expected to be positive if the model is appropriate, significant negative estimates of parameters would be taken as a clear indication that some other model is more appropriate. Thus, there are two criteria of the adequacy of a particular model. The first is the purely statistical criterion of whether the residuals are too large to be explained by chance alone. The second criterion is whether the estimates we obtain are consistent with biological theory. We shall see in due course that biologically nonsensical answers are often as clear a guide to the inadequacy of a model as the significance of residual effects.

(*c*) *The power of the tests of* G × E. The expected values of the normal deviate show that we would frequently expect to detect D_R, E_1 and E_2 as effects significantly different from zero, but that we would fail to detect $G \times E$ in a large number of cases. This statement may be quantified by calculating the power of each of the tests; that is, we can determine the proportion of studies of this size and structure in which we would reject at the 5 per cent level the four separate null hypotheses that D_R, $D_R E_2$, E_2 and E_1 were zero, given that the population parameters take the values used in generating the expected mean squares.

The power of the two-tailed test at the 5 per cent level is equal to the area under the normal curve having zero mean and unit variance between limits $1{\cdot}96 - c$ and infinity,

where c denotes the expected value of θ/σ_θ for a particular parameter. The values of $1{\cdot}96-c$ corresponding to the four parameters are $-1{\cdot}64$, $0{\cdot}27$, $-1{\cdot}18$, $-5{\cdot}34$ respectively, giving for the power of each of the tests 0·9495, 0·3936, 0·8810, 1·0000. That is, tests of the main effects of genes and environment are very powerful, but a study involving 1100 d.f. would only be 39 per cent certain of detecting $G \times E$ involving environmental differences between families, given that such interaction accounted for as much as 20 per cent of the total variance. Obviously, the power of the test would be substantially increased if the contribution of $G \times E$ were larger, but we have chosen to examine the consequences of a relatively small amount of $G \times E$ because that seems to be typical of real data examined to date.

The power calculations may be extended, following Eaves (1972), to determine the sample size which would be necessary to ensure greater power, say 95 per cent, for the detection of $G \times E$. For this purpose, we require the sample size which would produce an expected value of $1{\cdot}96+1{\cdot}65 = 3{\cdot}61$ for θ/σ_θ. We already know that the amount of $G \times E$ in our hypothetical population produces $\theta/\sigma_\theta = 1{\cdot}69$ for a total d.f. of 1100. The value of θ/σ_θ increases as the square root of the sample size, so writing k for the ratio of the required sample size to that used in our study we get $\sqrt{k} = 3{\cdot}61/1{\cdot}69$, giving $k = 4{\cdot}5629$. The sample size required to give 95 per cent certainty of detecting $G \times E$ at the 5 per cent level is thus $1100k$, or 5019. This fairly prohibitive requirement suggests that $G \times E$ which is not systematically related to genetical or environmental deviations (and therefore undetectable by other methods) is unlikely to be detected by fitting linear models to second-degree statistics unless the contribution of $G \times E$ is much larger in man than is usually the case in other organisms.

(*d*) *Assessing the bias in estimates.* If we proceed on the (false) assumption that $G \times E$ was absent biases would be introduced into the estimates of the main effects. The effect of these biases on subsequent inferences can be assessed accurately. We consider the simplest possibility which is likely to occur in practice, namely, the fitting of a model involving D_R, E_1 and E_2 to statistics derived from a population in which there is $G \times E$ due to $D_R E_2$ in addition to the three main effects.

Starting with the expected mean squares generated in the previous example, we attempted to fit three parameters, D_R, E_1 and E_2, by an iterative weighted least squares procedure, in which the expected values generated by fitting the model are used to produce new weights for each iteration. This is the procedure adopted in the analysis of actual data (e.g. Eaves & Eysenck, 1975).

The estimates are given by $(\mathbf{A'WA})^{-1}\mathbf{A'Wx}$ where $\mathbf{W}$ is evaluated for the parameter estimates which minimize the weighted squared deviations of 'observed' from predicted mean squares. The contribution of the 'true' parameters to these estimates may be obtained from the rows of $(\mathbf{A'WA})^{-1}\mathbf{A'WB}$ where $\mathbf{B}$ represents the 'true' model for the observed statistics.

For the case in which we fit only D_R, E_1 and E_2 to the population in which D_R, E_1, E_2 and $D_R E_2$ contribute to variation we find

$$\begin{pmatrix} \hat{D}_R \\ \hat{E}_2 \\ \hat{E}_1 \end{pmatrix} = \begin{pmatrix} 1 & \cdot & \cdot & 0{\cdot}4567 \\ \cdot & 1 & \cdot & 0{\cdot}2304 \\ \cdot & \cdot & 1 & 0{\cdot}0714 \end{pmatrix} \begin{pmatrix} D_R \\ E_2 \\ E_1 \\ D_R E_2 \end{pmatrix}.$$

As we might expect, the estimate of E_1 is virtually unbiased, having only $0{\cdot}0714 D_R E_2$ as bias. The variances of D_R and E_2, however, show a somewhat greater bias, D_R

containing $0{\cdot}4567D_R E_2$ and E_2 containing $0{\cdot}2304D_R E_2$. In this particular hypothetical example, therefore, the estimate of D_R is 176·10 compared with a true value of 135, and the estimate of E_2 is 76·99 compared with a true value of 56·25. Hence, $G \times E$ leads to a bias which overestimates both genetical and cultural effects if the contribution of $G \times E$ is ignored in fitting the model. The genetical component of the total variance is, however, $\frac{1}{2}D_R$ not D_R, and the bias in $\frac{1}{2}D_R$ is very nearly equal to that in E_2. That is, $G \times E$ biases equally our estimates of the contributions of genetical differences and environmental differences between families to the total variation.

Thus, although $G \times E$ may pass undetected if it is completely unsystematic, and although $G \times E$ will bias our estimates of genetical and environmental variance in such circumstances, it will none the less bias equally our estimates of D_R and E_2 in the same direction. In the light of this discussion, therefore, it seems appropriate to compute the estimate of the (narrow) heritability which would be obtained if a simple model were fitted in which the contribution of $G \times E$ was mistakenly ignored, and to compare the value with that computed from the true parameter values. We do this, not because we believe a knowledge of 'heritability' is fundamental to the problem of individual differences, but because certain authors (e.g. Moran, 1973; Layzer, 1974; Feldman & Lewontin, 1975) have focused much of their criticism of the analysis of human behaviour on estimates of 'heritability'. In writing an expression for the 'true' heritability, given $G \times E$, we include any variation due to $G \times E$ in the denominator only. Thus,

$$h^2 = \tfrac{1}{2}D_R/(\tfrac{1}{2}D_R + E_1 + E_2 + \tfrac{1}{2}D_R E_2) = 0{\cdot}30.$$

If, however, only D_R, E_1 and E_2 are fitted, then the heritability obtained is

$$h^2 = \tfrac{1}{2}\hat{D}_R/(\tfrac{1}{2}\hat{D}_R + \hat{E}_1 + \hat{E}_2) = 0{\cdot}3867.$$

That is, proceeding with the estimation of heritability on the assumption of no $G \times E$ is likely to lead to an overestimate of the true contribution of additive genetic variance to individual differences, when the estimate is based on the constellation of relatives we have considered here. Caution is advised, however, since this is not a general rule. Genotype–environmental interaction could lead to underestimation of additive genetical variance from transgenerational data (e.g. parent–offspring covariances), and the precise nature of any bias would have to be determined for each specific situation. Indeed, one drawback of all calculations of this kind is their lack of generality. We can merely explore the weaknesses of particular studies under a selected range of circumstances.

Although we sometimes include $G \times E$ with the non-heritable agencies, for predictive purposes we should recognize that the existence of $G \times E$ indicates that sensitivity to the environment is to a greater or lesser degree under genetical control. This enables expectations for $G \times E$ to be written in terms of gene effects. Furthermore, the existence of $G \times E$ indicates that sensitivity to the environment is potentially subject to the influence of selection. To affirm, as many have done, that the certainty of genotype—environmental interactions undermines genetical analysis, is both to exaggerate their significance relative to the overall variation in a population and to miss their potential biological significance.

8.2. *Genotype–environmental covariance*

Last (1976) conducted similar studies in relation to many other effects, including those of assortative mating, dominance and genotype–environment covariance. We consider some of her results relating to the detection of genotype–environment covariance.

(*a*) *Sibling effects.* Using the model developed by Eaves (1976*a*), Last investigated the detection of environmental variation and genotype–environmental covariation due to the effect of one sibling on another. A simple system was considered in which all gene effects were additive and in which the only common environmental component was that due to the genetical covariance between members of a twin (or sibling) pair. The total variance of singletons was thus specified as $\sigma_s^2 = \frac{1}{2}D_R + E_1$, there being no variation due to sibling effects in individuals reared alone. Following the procedure previously adopted, E_1 was equated to $\frac{1}{2}D_R$ and σ_s^2 was arbitrarily fixed at 225, giving $E_1 = 112{\cdot}5$ and $D_R = 225$. In order to specify the sibling effects, we have to provide values for $D_{\ddot{R}}$ (the 'genetic environmental' variance due to sibling effects) and $D_{\dot{R}}$ (the genotype–environmental covariance parameter).

Table 8. A simplified model for sibling effects used in simulation

	Parameter			
Mean square	D_R	$D_{\ddot{R}}$	$D_{\dot{R}}$	E_1
Between MZT	1	1	2	1
Within MZT	·	·	·	1
Between DZT	$\frac{3}{4}$	$\frac{3}{4}$	$1\frac{1}{2}$	1
Within DZT	$\frac{1}{4}$	$\frac{1}{4}$	$-\frac{1}{2}$	1
Between MZA	1	·	·	1
Within MZA	·	·	·	1
Between DZA	$\frac{3}{4}$	·	·	1
Within DZA	$\frac{1}{4}$	·	·	1
Between UT	$\frac{1}{2}$	$\frac{1}{2}$	1	1
Within UT	$\frac{1}{2}$	$\frac{1}{2}$	−1	1
Singletons	$\frac{1}{2}$	$\frac{1}{2}$	1	1

The model, given in Table 8, shows that the contributions of environmental variance, specified by $D_{\ddot{R}}$, and of Cov GE, specified by $D_{\dot{R}}$, depend on the degree of relationship between members of the pairs contributing to a particular statistic. Thus, although there is an explicit interpretation of D_R, $D_{\ddot{R}}$ and $D_{\dot{R}}$, their relative contributions to particular statistics change. This means, further, that any estimate of ρ_{ge} depends on the degree of relationship, and on the relative contribution of E_1 and $D_{\ddot{R}}$ to the total variance. It is small wonder, therefore, that previous attempts to specify Cov GE have foundered at the outset.

For the purposes of simulation, Last assumed that $D_{\ddot{R}} = \frac{1}{2}D_R$; that is, the environmental variance due to sibling effects was assumed to be half that due to E_1. The only remaining difficulty was the specification of $D_{\dot{R}}$. This parameter is, in effect, a measure of the degree of pleiotropy or association between the direct effects of genes on an individual's phenotype and their indirect effect, through the environment, on the performance of a sibling. In this sense, and in this sense only, it is meaningful to define, and solely for the purpose of these simulations, that $\rho_{ge} = D_{\dot{R}}/\sqrt{(D_R D_{\ddot{R}})}$.

When the allele frequencies are equal at every locus contributing either to D_R or $D_{\ddot{R}}$, and when all loci have consistent and equal effects (see Eaves & Gale, 1974, for a definition of consistency in this context), then ρ_{ge} is the proportion of loci which contribut both to D_R and $D_{\ddot{R}}$. That is, ρ_{ge} is the proportion of loci having both direct effects on the phenotype, and indirect effects on the phenotype of siblings. Last chose to work

with $\rho_{ge} = 0{\cdot}5$. The values employed in generating her statistics, therefore, were $D_{\ddot{R}} = 112{\cdot}5$ and $D_{\dot{R}} = 79{\cdot}5495$.

With the sample structure she had assumed (1100 d.f.), Last was able to derive values for the powers of the tests of the four parameters, D_R, $D_{\dot{R}}$, $D_{\ddot{R}}$ and E_1. These values were 1·00, 0·99, 0·67 and 1·00 respectively. Thus, we see that the present design is quite powerful for detecting sibling effects, especially when these covary with the direct effects of genetical differences. The fact that the test of $D_{\dot{R}}$ is more powerful than that of $D_{\ddot{R}}$ arises because the test of $D_{\ddot{R}}$ depends merely on a comparison between the total variances of individuals reared as pairs and singletons, whereas the significance of $D_{\dot{R}}$ depends on a very precise pattern of expectations which involve the individual mean squares as well as the total variances.

By considering the case in which $D_{\dot{R}}$ is positive we have assumed a cooperative situation in which a high performance of one sibling enhances the performance of another. We must also consider the possibility of a competitive situation in which the success of one sibling is achieved at the cost of the other. Last simulated competition by using the same parameters as the cooperative model, except that $D_{\dot{R}}$ was replaced by a negative value, $-79{\cdot}5495$. The values obtained for the powers of the four tests of D_R, $D_{\dot{R}}$, $D_{\ddot{R}}$ and E_1 were now 1·00, 0·95, 0·57 and 1·00, suggesting that the detection of competitive effects is no less viable than the detection of cooperation. Indeed, we suspect that competition may be easier to detect than cooperation in experimental designs which lack individuals reared as singletons.

(*b*) *Cultural transmission.* Finally, we examine the detection of genotype–environment covariation arising as a result of the cultural impact of parents on their offspring. Eaves (1976*b*) has shown how the empirical components E_2 and $\mathrm{Cov}\, g_2 e_2$ may be represented in terms of a cultural transmission model in which the phenotypic variation between parents is perpetuated culturally as well as genetically in the offspring generation. Using b to denote the regression of the family environment of an offspring on the phenotype of his parents, it can be shown, in a randomly mating population that

$$E_2 = \left(\frac{2b^2}{1-2b^2}\right)\left[\left(\frac{1+b}{1-b}\right)\tfrac{1}{2}D_R + E_1\right]$$

and

$$2\,\mathrm{Cov}\, g_2 e_2 = \frac{2b}{1-b}(\tfrac{1}{2}D_R) \quad \text{for } |b| < \tfrac{1}{2}.$$

Any common environmental effects of a non-cultural kind are excluded from this formulation of the model, but need not be excluded in principle. The magnitude of E_2 and $\mathrm{Cov}\, g_2 e_2$ depend on the size of $\frac{1}{2}D_R$, but E_2 is further increased by the contribution of dominance and random environmental factors. We consider one of the simulation studies conducted by Last (1976) which assumed that, in the absence of cultural transmission and given complete additivity, $\frac{1}{2}D_R = E_1 = 112{\cdot}5$, i.e. that the heritability of the trait would be 0·5 in the absence of cultural effects. Assuming that cultural transmission was entirely responsible for any common environmental effects, it is possible to calculate expected values of E_2 and $2\,\mathrm{Cov}\, g_2 e_2$ for a given b. Last considered the consequences of $b = 0{\cdot}1$ and $b = 0{\cdot}25$. For $b = 0{\cdot}1$, we get $E_2 = 6{\cdot}76$ and $2\,\mathrm{Cov}\, g_2 e_2 = 50{\cdot}00$. These values may be compared with those of 61·61 and 150·00 for E_2 and $2\,\mathrm{Cov}\, g_2 e_2$ when $b = 0{\cdot}25$, given random mating, additive gene action and $\frac{1}{2}D_R/(\frac{1}{2}D_R + E_1) = 0{\cdot}5$.

Strictly speaking, since only one new parameter, b, is required to specify the consequences of cultural transmission for E_2 and $\mathrm{Cov}\, g_2 e_2$ it would be appropriate to employ the more powerful approach of non-linear weighted least squares (e.g. Eaves, 1975) to estimate b, D_R and E_1, but we choose to continue the approach adopted so far and estimate four parameters E_2, $\mathrm{Cov}\, g_2 e_2$, E_1 and D_R. This is sufficient for our purpose, but does not enforce the constraint upon E_2 and $\mathrm{Cov}\, g_2 e_2$ which is implied in the model of cultural transmission. We may find, however, that real data do not fit exactly into the mould of our equilibrium model, so the restraint may need to be relaxed anyway in practice. From this point of view, the approach we adopt here may be more generally applicable.

For the case of $b = 0{\cdot}1$, the powers of the tests of the parameters were found to be 1·00, 0·39, 0·76 and 1·00 for D_R, E_2, $\mathrm{Cov}\, g_2 e_2$ and E_1 respectively. As b or h^2 increase, of course, the tests become much more powerful. Thus, for $b = 0{\cdot}25$ the powers become 1·00, 0·99, 0·99 and 1·00. If the contribution of $\frac{1}{2}D_R$ is increased relative to that of E_1, so that $\frac{1}{2}D_R/\frac{1}{2}D_R + E_1 = 0{\cdot}9$ (i.e. $D_R = 405$ and $E_1 = 11{\cdot}25$), then for $b = 0{\cdot}1$, we obtain $E_2 = 10{\cdot}33$ and $2\,\mathrm{Cov}\, g_2 e_2 = 90$. For $b = 0{\cdot}25$, we obtain $E_2 = 98{\cdot}04$ and $2\,\mathrm{Cov}\, g_2 e_2 = 270$. The power of the test for E_2 is 0·87, when $b = 0{\cdot}1$, and 1·00 when $b = 0{\cdot}25$. The powers of the tests of genotype–environmental covariance are 0·93 and 1·00 for $b = 0{\cdot}1$ and 0·25, respectively.

We give somewhat fuller results for the cultural transmission model because of the potential theoretical significance of $\mathrm{Cov}\, g_2 e_2$. The power of the test of $\mathrm{Cov}\, g_2 e_2$ is found to be reasonable ($> 0{\cdot}75$) for this sample structure when $b \geqslant 0{\cdot}1$ and when $\frac{1}{2}D_R \geqslant E_1$. In terms of the more usual models for genotype–environmental correlation, these parameter values coincide with overall genotype–environmental correlations of 0·31 (for $b = 0{\cdot}1$, $\frac{1}{2}D_R = E_1$) and 0·65 (for $b = 0{\cdot}25$, $\frac{1}{2}D_R = E_1$).

We can obtain the linear combination of the observed statistics which yields the weighted least squares estimate of $2\,\mathrm{Cov}\, g_2 e_2$. For the case of $b = 0{\cdot}25$ and $h^2 = 0{\cdot}5$, the coefficients of the 11 statistics in the estimator of $2\,\mathrm{Cov}\, g_2 e_2$ are given in Table 9.

Table 9. Contributions of 11 statistics to weighted least squares estimate of genotype–environmental covariance

Mean square	Coefficient
Between MZT	0·1616
Within MZT	0·4507
Between DZT	0·1974
Within DZT	−0·1048
Between MZA	−0·1209
Within MZA	−0·2304
Between DZA	−0·1395
Within DZA	−0·1964
Between UT	−0·1564
Within UT	−0·1433
Singletons	0·2821

It will be seen that the test of $\mathrm{Cov}\, g_2 e_2$ largely depends on the comparison of the total variance of individuals reared by their natural parents with that of individuals reared by foster parents. This observation is important for two reasons. Firstly, it does not really matter whether the fostered individuals are twins or not, nor even if they are reared in pairs. Secondly, it is a serious waste of information, and produces a grave

risk of unsuspected bias, to attempt the detection of Cov GE from a comparison of correlation coefficients rather than mean squares. Whatever compromises may be forced upon us by the inadequacies of published data summaries, the analysis of correlations as a matter of policy cannot be defended on biological, psychological or statistical grounds. Any advantage the correlation coefficient may have as a statistic of compelling simplicity for the purpose of communication is rapidly lost in any serious attempt to analyse the causes of individual differences.

In view of the discussion in the literature about the contribution of dominance and Cov $g_2 e_2$ to individual differences in IQ, it seems appropriate to investigate the consequences of mistakenly fitting a dominance parameter instead of Cov $g_2 e_2$ to data in which that covariance is known to be present. Last found that fitting a D_R, H_R, E_1, E_2 model to the data above ($b = 0{\cdot}1$, $\frac{1}{2}D_R = E_1$) gave a negative and non-significant dominance parameter. If we assume that the residuals all take their expected values apart from the effect of Cov $g_2 e_2$, we would expect the χ^2 test of goodness of fit to be 13·42 for 7 d.f., indicating borderline failure of the model. This evidence, together with the negative value for the dominance variation, would lead us to seek an alternative explanation for the pattern of variation. Thus, it would seem that we are unlikely in practice to confuse positive Cov $g_2 e_2$ with dominance. We suspect, however, that it might be possible to confuse negative genotype–environmental covariance with dominance in borderline cases (including cases of competition), since some of the consequences of negative genotype–environmental covariance may reproduce certain of the effects of dominance.

9. Discussion

There have been many attempts to cast doubt on the legitimacy and value of studies of human differences. Most criticisms are purely negative and many lack any systematic quantitative framework which could form the basis for future research. Indeed, certain critics seem to have little commitment to anything more than an *ad hoc* explanation of a particular set of data. The quality of criticism can be seen in recent attempts to reinterpret twin data on the basis of a 'pseudo-environmental' model (see e.g. Schwartz & Schwartz, 1974). The starting point for such an explanation of twin differences is the assumption that the environmental factors which affect the development of monozygotic twins are more alike within pairs than those which affect dizygotic twins. The strength of such explanations, from the viewpoint of their adherents, is that they are imprecise and do not propose any general mechanism. Consequently, parameters can be modified at will to obtain an explanation which is consistent with any set of twin data. From the scientific viewpoint, however, this is also their weakness. A theory which is sufficiently plastic to explain everything after the event cannot form a useful basis for predictive statements.

The long-term credibility and utility of a particular theory of individual differences cannot rest merely on an explanation of twin differences (or, for that matter, on the explanation of any other restricted set of data), but must rest on the ability of the theory to encompass a wide variety of data with the minimum of special pleading. Such theories have been labelled by Urbach (1974) as 'progressive' in contrast to 'degenerating' theories, which appeal constantly to *ad hoc* explanations. In our view, much misunderstanding and controversy stems from a failure to relate individual statistics to a wider predictive framework. This is particularly the case with regard to the discussion of genotype–environmental covariation (see e.g. Jensen, 1975; and,

in criticism, Goldberger & Lewontin, 1976), which has often lacked the overview of a theoretical approach which transcends immediate explanations of one or two statistics.

The advocates of *ad hoc* explanations of twin differences, for example, will command little credibility until they can offer a serious, systematic, theoretical extension of their approach in a form which enables successful predictions to be made. Such a theoretical extension would have to be quantitative, and would require some strong precise conjecture about the mechanism responsible for individual differences. So far, those who have criticized genotype–environmental theories have advanced no alternative general mechanism for the maintenance of individual differences, and consequently have never considered some of the implications of such mechanisms in detail, even for the restricted instance of twin data. Indeed, some of the few serious steps in this direction have been taken by geneticists (e.g. Cavilli-Sforza & Feldman, 1973; Eaves, 1976*a*, *b*). Thus, for example, Eaves (1976*a*) has shown how one theory of environmental differences in twins can be given a precise quantitative formulation and leads us to expect, among other things, differences in the total variances between MZ and DZ twins. Such differences would falsify most simple genotype–environmental models, and also form a valuable point of departure for a study of the nature of environmental variation, since the theory can be used to predict the findings for other kinds of family. A recent paper (Zajonc, 1976) uses a somewhat different approach to essentially the same problem with reference to the environmental factors which influence cognitive development. It is only when a theory receives such quantitative formulation that it can become a serious competitor in a quantitative science.

We have tried in this paper to supply some of the missing framework in the hope that subsequent discussion will be better informed. Except where it is absolutely unavoidable we have chosen the approach of simulation rather than the analysis of actual data, because we have seen how easy it is for matters of method and principle to be buried under arguments about data whose quality leaves much to be desired. By considering such problems as experimental design and power, we have tried to show that the presence of non-additivity and non-independence does not constitute a fundamental barrier to quantitative analysis. Each case has to be considered on its own merits. The presence of $G \times E$, for example, does not necessarily lead to errors of inference. The larger the effect of $G \times E$, the more likely it is to be detected. The smaller the effect, the smaller the error which stems from failure of detection. Arguments which have often raged in a simple-minded qualitative manner in the literature, are in fact quantitative. Such qualitative arguments are basically unhelpful to anyone interested in the truth about individual differences for a quantitative trait. No verbal argument can detract from the proven success of close quantitative argument and statistical reasoning in dealing with continuous variation in any organism, human or otherwise. Genotype–environment interaction and genotype–environment covariation are not hidden mysteries, they are potentially detectable realities which are of the utmost importance for our understanding of human differences. The negative and frequently ill-formed criticisms by a great many authoritative writers with little practical experience in this area, merely spread confusion and inhibit scientific progress.

Acknowledgements

We are indebted to Dr M. J. Kearsey for allowing us to use his unpublished simulation program. The computations were conducted on the Birmingham University ICL 1906A computer. We thank Judi Pritchard for her patient secretarial help, and our many colleagues and collaborators for access to data. The work is part of a research programme in psychogenetics supported by the Medical Research Council. Krystyna Last currently holds an MRC Research Studentship.

References

Bock, R. D. & Kolakowski, D. (1973). Further evidence of sex-linked major gene influence in twins' spatial visualizing ability. *Am. J. hum. Genet.* **25**, 1–14.

Breese, E. L. & Mather, K. (1957). The organisation of polygenic activity within a chromosome of Drosophila. I. Hair characters. *Heredity* **11**, 373–395.

Breese, E. L. & Mather, K. (1960). The organisation of polygenic activity within a chromosome in Drosophila. II. Viability. *Heredity* **14**, 375–399.

Broadhurst, P. L. & Jinks, J. L. (1974). What genetical architecture can tell us about the natural selection of behavioural traits. In Van Abeelen (ed.), *The Genetics of Behaviour*. Amsterdam: North Holland.

Brumpton, R. J. (1973). Studies on the control of chiasma formation and on the expression of developmental characters. Unpublished PhD thesis, University of Birmingham, England.

Bucio-Alanis, L., Perkins, J. & Jinks, J. L. (1969). Environmental and genotype–environmental components of variability. V. Segregating generations. *Heredity* **24**, 115–127.

Burt, C. (1966). The genetic determination of intelligence: A study of monozygotic twins reared together and apart. *Br. J. Psychol.* **57**, 137–153.

Cattell, R. B. (1960). The multiple abstract variance analysis equations, and solutions for nature: nurture research on continuous variables. *Psychol. Rev.* **67**, 353–372.

Cavalli-Sforza, L. & Feldman, M. (1973). Cultural versus biological inheritance: Phenotypic transmission from parents to children (a theory of the effect of parental phenotypes on children's phenotypes). *Am. J. hum. Genet.* **25**, 618–637.

Comstock, R. E. & Robinson, H. F. (1952). Estimation of average dominance of genes. In W. J. Gowan (ed.), *Heterosis*. Ames, Iowa: Iowa State College Press.

DeFries, J. C. *et al.* (1976). Parent–offspring resemblance for specific cognitive abilities in two ethnic groups. *Nature, Lond.* **261**, 131–133.

Eaves, L. J. (1969). The genetic analysis of continuous variation: A comparison of experimental designs applicable to human data. *Br. J. math. statist. Psychol.* **22**, 131–147.

Eaves, L. J. (1970). Aspects of human psychogenetics. Unpublished PhD thesis, University of Birmingham, England.

Eaves, L. J. (1972). Computer simulation of sample size and experimental design in human psychogenetics. *Psychol. Bull.* **77**, 144–152.

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

Eaves, L. J. (1975). Testing models for variation in intelligence. *Heredity* **34**, 132–136.

Eaves, L. J. (1976*a*). A model for sibling effects in man. *Heredity* **36**, 205–214.

Eaves, L. J. (1976*b*). The effect of cultural transmission on continuous variation. *Heredity* **37**, 41–57.

Eaves, L. J. & Eysenck, H. J. (1975). The nature of extraversion: A genetical analysis. *J. Person. soc. Psychol.* **32**, 102–112.

Eaves, L. J. & Eysenck, H. J. (1977). A genotype–environmental model for psychoticism. *Adv. Behav. Res. Ther.* I (in press).

Eaves, L. J. & Gale, J. S. (1974). A method for analysing the genetic basis of covariation. *Behav. Genet.* **4**, 253–267.

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

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433.

Fisher, R. A., Immer, F. R. & Tedin, O. (1932). The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. *Genetics* **17**, 107–124.

Fripp, Y. J. & Caten, C. E. (1973). Genotype–environmental interactions in *Schizophyllum commune*. III. The relationship between mean expression and sensitivity to change in environment. *Heredity* **30**, 341–349.

Goldberger, A. S. & Lewontin, R. C. (1976). Jensen's Twin Fantasy. Discussion Paper No. 341. University of Wisconsin, Madison: Institute for Research on Poverty.

Haldane, J. B. S. (1946). The interaction of nature and nurture. *Ann. Eugen.* **13**, 197–205.

Hanson, W. D. & Robinson, H. F. (eds). (1963). *Statistical Genetics and Plant Breeding*. Washington, D.C.: National Academy of Sciences–National Research Council.

Hogarth, R. (1974). Monozygotic and dizygotic twins reared together: Sensitivity and heritability estimates. *Br. J. math. statist. Psychol.* **27**, 1–13.

Jencks, C. *et al.* (1973). *Inequality: A Reassessment of Effect of Family and Schooling in America*. New York: Basic Books.

Jensen, A. R. (1975). The meaning of heritability in the behavioural sciences. *Educ. Psychol.* **11**, 171–183.
Jinks, J. L. (1955). A survey of the genetical basis of heterosis in a variety of diallel crosses. *Heredity* **9**, 223–238.
Jinks, J. L. & Broadhurst, P. L. (1963). Diallel analysis of litter size and body weight in rats. *Heredity* **18**, 319–336.
Jinks, J. L. & Connolly, V. (1975). Determination of the environmental sensitivity of selection lines by the selection environment. *Heredity* **34**, 401–406.
Jinks, J. L. & Eaves, L. J. (1974). IQ and inequality. *Nature, Lond.* **248**, 287–289.
Jinks, J. L. & Fulker, D. W. (1970). A comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychol. Bull.* **73**, 311–349.
Jinks, J. L. & Jones, R. M. (1958). Estimation of the components of heterosis. *Genetics* **43**, 223–234.
Jinks, J. L. & Perkins, J. M. (1969). The detection of linked epistatic genes for a metrical trait. *Heredity* **24**, 465–475.
Jinks, J. L. & Perkins, J. M. (1970*a*). A general method for the detection of additive, dominance and epistatic components of variation. III. F_2 and backcross populations. *Heredity* **25**, 419–429.
Jinks, J. L. & Perkins, J. M. (1970*b*). Environmental and genotype–environmental components of variability. VII. Simultaneous prediction across environments and generations. *Heredity* **25**, 475–480.
Kamin, L. (1974). *The Science and Politics of IQ*. Potomac, Md: Erlbaum.
Kempthorne, O. (1960). *Biometrical Genetics*. London: Pergamon Press.
Last, Krystyna (1976). Genetical aspects of human behaviour. Unpublished dissertation, Department of Genetics, University of Birmingham, England.
Layzer, D. (1974). Heritability analysis of IQ scores: Science or numerology? *Science, N.Y.* **183**, 1259–1266.
Lewontin, R. C. (1974). The analysis of variance and the analysis of causes. *Am. J. hum. Genet.* **26**, 400–411.
Lewontin, R. C. (1976). Race differences in intelligence (review of Loehlin, J. C., Lindzey, G. & Spuhler, J. N. *Race Differences in Intelligence*. San Francisco: W. H. Freeman). *Am. J. hum. Genet.* **28**, 92–97.
Loehlin, J. C. (1965). Some methodological problems in Cattell's multiple abstract variance analysis. *Psychol. Rev.* **72**, 156–161.
Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Reading, Mass: Addison-Wesley.
Martin, N. G. (1976). Aspects of human behaviour genetics. Unpublished dissertation, Department of Genetics, University of Birmingham, England.
Martin, N. G. & Eysenck, H. J. (1976). Genetic factors in sexual behaviour. In H. J. Eysenck, *Sex and Personality*. London: Open Books.
Mather, K. (1943). Polygenic inheritance and natural selection. *Biol. Revs. Camb. phil. Soc.* **18**, 32–64.
Mather, K. (1967). Complementary and duplicate gene interaction in biometrical genetics. *Heredity* **22**, 97–103.
Mather, K. (1973). *Genetical Structure of Populations*. London: Chapman & Hall.
Mather, K. (1974). Non-allelic interactions in continuous variation of randomly breeding populations. *Heredity* **32**, 414–419.
Mather, K. & Caligari, P. D. S. (1975). Genotype–environment interaction. III. Interactions in *Drosophila melanogaster*. *Proc. R. Soc. Lond.* B **191**, 387–411.
Mather, K. & Jinks, J. L. (1971). *Biometrical Genetics: The Study of Continuous Variation*. London: Chapman & Hall.
Mather, K. & Jones, R. M. (1958). Interaction of genotype and environment in continuous variation. I. Description. *Biometrics* **14**, 343–359.
Moran, P. A. P. (1973). A note on heritability and the correlation between relatives. *Ann. hum. Genet.* **37**, 217.
Morton, N. E. (1974). Analysis of family resemblance. I. Introduction. *Am. J. hum. Genet.* **26**, 318–330.
Perkins, J. M. & Jinks, J. L. (1968). Environmental and genotype–environmental components of variability. III. Multiple lines and crosses. *Heredity* **23**, 339–356.
Perkins, J. M. & Jinks, J. L. (1973). The assessment and specificity of environmental and genotype–environmental components of variability. *Heredity* **30**, 111–126.
Schwartz, M. & Schwartz, J. (1974). Evidence against a genetical component to performance on IQ tests. *Nature, Lond.* **248**, 84–85.

Thomas, H. (1975). A model for interpreting genetic correlations with estimates of parameters. *Psychol. Bull.* **52**, 711–719.
Urbach, P. (1974). Progress and degeneration in the 'IQ Debate' I and II. *Br. J. phil. Sci.* **25**, 99–135; 235–259.
Vandenberg, S. G. (1962). The hereditary abilities study: Hereditary components in a psychological test battery. *Am. J. hum. Genet.* **14**, 220–237.
Vetta, A. (1976). Correction to Fischer's correlation between relatives and environmental effects. *Nature, Lond.* **263**, 316–317.
Vetta, A. & Smith, C. A. B. (1974). Comments on Fisher's theory of assortative mating. *Ann. hum. Genet.* **38**, 243–248.
Wilson, S. R. (1973). The correlation between relatives under the multifactorial model with assortative mating. I. The multifactorial model with assortative mating. *Ann. hum. Genet.* **37**, 189–204.
Zajonc, R. B. (1976). Family configuration and intelligence. *Science, N.Y.* **192**, 227–236.

Received 14 June 1976; revised version received 11 August 1976.

Requests for reprints should be addressed to L. J. Eaves, Department of Genetics, University of Birmingham, B15 2TT, England.