GENETICAL ASPECTS OF HUMAN BEHAVIOUR

A thesis presented to the

Faculty of Science and Engineering

of the

University of Birmingham

by

Krystyna Last

for the degree of

Doctor of Philosophy

Department of Genetics

University of Birmingham

January, 1978

SYNOPSIS

Race and sex differences in mean ability follow the usual pattern in Black and White twins from the US. Whites are more variable, partly due to greater differences between families. This may reflect greater cultural heterogeneity within populations or greater heterogeneity of the White population. Differences in mean performance of Whites from different regions partly explains this heterogeneity. Many apparent interactions between genotypic and environmental factors can be attributed to test construction.

A twin study confirms the role of common and specific environmental factors, assortative mating and additive genetical variation in determining social attitudes. Path models, specifying the effects of cultural transmission and the mating system, are fitted to twin and adoption data using the maximum likelihood methods of pedigree analysis. The between families component is apparently explained by the effects of the mating system in these data, but a more powerful test is needed.

Simulation studies show that exclusion of potentially important factors (e.g. genotype-environment interaction) from models of variation does not necessarily lead to overestimation of genetical effects, as commonly supposed. The precise biases depend on experimental design, population parameter values and the actual models fitted to the data, but can be quantified exactly and probable errors of inference can be estimated.

ACKNOWLEDGEMENTS

The people who have contributed to the completion of this thesis are too numerous to mention and must include all my colleagues and friends. But my especial gratitude goes to:

Dr. Lindon Eaves,

for his help and advice throughout. His considerable expertise and experience coupled with his great enthusiasm have proved invaluable. I may truly say that his encouragement and friendship have made this work possible and have made my studentship an enjoyable and valuable experience.

Discussion with colleagues has proved useful and stimulating. I want particularly to mention the liveliness and help of my fellow research students, Dr. N.G. Martin, Mr. P.A. Young and Miss M.P. Towey, and the advice of Professor J.L. Jinks, Dr. M.J. Kearsey and Dr. J.S. Gale.

The work would not have been possible without the data provided by Professor H.J. Eysenk from the Institute of Psychiatry, University of London and Dr. R.T. Osborne, Professor of Psychology at the University of Georgia, U.S.A., whose collaboration and friendship I have enjoyed throughout.

The preparation of the scripts has involved a number of helpful and efficient typists. My thanks goes to:

Mrs. Jean Bourne, for the typing and layout of the text

Mrs. Judy Barker, for the Tables of Part A

Mrs. J. Moss, for the Tables of Part B

Miss Bardell, for the Tables of Part C.

I wish to express my gratitude to Mr. D.R. Kelly for his cheerful help in the gruelling task of compiling the tables in Appendix G and checking the script and tables.

The University Computer Centre Staff have given advice on a number of technical points. Mr. Roy Pearce has been particularly helpful in making successful communication with the University of Manchester's Computer Centre possible.

Financial support for the duration of my studentship was provided by the Medical Research Council.

LIST OF CONTENTS

SECTION 3: THE SCALE OF MEASUREMENT

SECTION 4: AGE EFFECTS IN THESE DATA

SECTION 5: PRELIMINARY ANALYSIS

LIST OF TABLES

PART A - THE GEORGIA TWIN STUDY

SECTION 1: A DESCRIPTION OF THE DATA

SECTION 2: PRELIMINARY ANALYSIS OF THE DATA

SECTION 7: ANALYSIS OF A "GENERAL FACTOR"

PART B - A STUDY OF SOCIAL ATTITUDES

SECTION 1: DESCRIPTION OF THE STUDY

SECTION 2: THE STRUCTURE OF ATTITUDES IN THESE DATA

SECTION 7: FITTING MODELS SIMULTANEOUSLY TO THE TWIN DATA AND THE UNBALANCED PEDIGREES OF THE ADOPTION STUDY

SECTION 8: AN ALTERNATIVE APPROACH THROUGH THE USE OF PATH MODELS

PART C - A STUDY OF POWER AND BIASES IN HUMAN BEHAVIOUR GENETICS

LIST OF FIGURES

PART A - THE GEORGIA TWIN STUDY

PART B - A STUDY OF SOCIAL ATTITUDES

GENERAL INTRODUCTION

This thesis is divided into three distinct parts which at first glance may appear unrelated. However, they are linked through a common philosophy and aim and show progress in the concepts and theory of behavioural genetics, methodological advances and the direction of ideas about data collection.

In all three parts model-building and hypothesis-testing are of paramount importance and the underlying philosophy may be summarised by the enumeration of five distinct stages of data analysis:

1. Construction of a series of alternative models for the observed pattern of variation
2. A clear statement of the assumptions on which these alternative hypotheses about the data are based
3. Tests of the assumptions whenever possible
4. Efficient estimation of parameters of the models
5. Powerful tests of the adequacy of the models using suitable statistical criteria.

Failure to follow the guidelines implied by these five points will in many instances lead to errors of inference. The overall aim of this thesis may, thus, be stated as the attempt, throughout the analyses, to recognise when errors of inference may occur and to show precisely how biases produced by the failure to test critical assumptions can affect our conclusions. This is the particular concern of Part C.

The earliest work in behavioural genetics largely concerned the determination of ability (e.g. Galton, 1869; Burt, 1912; Burks, 1928)

and the twin design suggested by Galton (1869) has commonly been used. The work in Part A of this thesis continues this tradition and considers the causes of variation in a number of measures of ability in Black and White twins from the southern United States. It allows a detailed and thorough analysis of data in an area where poor data, unrealistic analyses and mistaken interpretations have caused much confusion. However, the models tested are not original and are well known to biometrical-geneticists (see Jinks and Fulker, 1970), although they have not commonly been employed in studies of Black and White twins. It was hoped that the rigour of the analyses in Part A would make powerful influences possible and minimise the possibility of mistaken conclusions.

Although all types of relative provide information about genetical and environmental components of variation, most commonly monozygotic and dizygotic twins have been used (as in the Georgia Twin Study of Part A), since generally these were thought to provide the most powerful initial tests of simple genotype-environmental models. It was recognised that, after preliminary analysis, other experimental designs may be more useful. For example, parent-offspring covariances provide a powerful test for sex-linkage. In view of the power calculations in Parts A and C, we suggest that alternative experimental designs should be considered from the outset.

The weakness of two statistics commonly used by behavioural geneticists in the past are demonstrated in Part A. Many previous workers have used the intraclass correlation coefficient as a measure of the resemblance between pairs of relatives, rather than analysing variance components (e.g. Loehlin and Nichols, 1976). This approach is useful when only straightforward analysis is performed and the causes of

variation are very simple. However, with behavioural traits, this is usually not the case and we may be seriously misled by using the correlational approach. Using correlations rather than variance components assumes equality of the total variances between groups. Therefore, the assumption is made that all effects leading to inequality of the total variances are absent. If this is not the case then estimates obtained from the data will be biased, and information will be lost about effects leading to unequal variances. Jinks and Fulker (1970) have enumerated several factors producing inequality of the total variances:

1. The genetical components differ between groups.
2. Between and within pairs environmental components differ between groups.
3. Genotype-environment covariance is present.

Recently, Eaves (1976a) showed how competition/cooperation between siblings leads to genotype-environment covariance which produces a predictable pattern of inequality in the total variances for monozygotic and dizygotic twins and unrelated individuals reared together.

Thus, if we test the total variances for equality before using correlations, we can ensure that we are not mislead into accepting a simple model of variation where a more complex model is the true explanation. However, since the analysis that can be performed using correlation coefficients is limited and there are problems when the total variances are unequal, it would be better and easier to make maximum use of the information available in the data and to base all analysis of this type of data on the biometrical-genetical approach as applied to variance components.

The limitations of the simple analyses which may be undertaken using the correlational approach are amply demonstrated in Part A using the data of the Georgia Twin Study.

The second statistic which we show to be of limited value in behavioural genetics is the heritability. This is often used as a summary statistic and is the amount of genetical variation expressed as a proportion of the total variation. It gives the relative contributions of genetical and environmental components within any population.

Several authors (e.g. Newman, Freeman and Holzinger, 1937; Nichols, 1965; Vandenberg, 1966) have used data on MZ and DZ twins to generate so-called "heritability ratios". Such ratios, however, bear no simple relationships to either the broad or narrow heritabilities familier to quantitative geneticists (Jinks and Fulker, 1970) unless some radical simplifying assumptions are made. The approaches generally used offer no test of these assumptions. The model fitting approach of biometrical genetics, however, combines parameter estimation and tests of assumptions in the same analysis with the result that many fundamental biases and inconsistencies of older approaches can be avoided.

The heritability ratio may not be used for comparisons between groups. Two populations, where the environmental and genetical components of variation are quite different may have similar heritabilities by chance. The analyses presented in Part A provide conclusive evidence on this point and reinforce our belief that analysis of variance components provides the most efficient use of data, when its structure is regular as in the case of twins.

The problem of genotype-environment interaction for measures of ability and the extent to which apparent interactions between within-

families environmental differences and genetical differences can be attributed simply to the scale of measurement are considered. A second cause of non-independence of genotype and environment i.e. genotype-environment covariation is not considered, since previous workers have found no substantial effect of CovGE in the determination of ability in human populations (e.g. Jinks and Fulker, 1970).

Previous work suggests a fairly simple genotype-environmental mechanism for individual differences in intelligence. The pattern of variation is that of a trait determined by environmental differences within families and genetical differences. There is no indication that environmental differences between families produce any substantial effect on intelligence. However, additive genetical variation produced by assortative mating (Burt and Howard, 1956; Jinks and Fulker, 1970) and dominance variation (Schull and Neel, 1965; Spuhler, 1967; Jinks and Fulker, 1970) are implicated. Both these phenomena show that ability as measured by conventional intelligence tests is a biologically meaningful variable and, in particular, the presence of strong directional dominance for I.Q. indicates that during man's evolution high I.Q. contributed to reproductive fitness. Thus intelligence might appear a suitable trait for the study of biological and cultural evolution. However, this may be precluded in the latter case by recent work of Eaves (1976a, b) which postulates three mechanisms which could lead to genotype-environment covariation. His specification of one of these mechanisms in algebraic terms provides a theoretical model for cultural transmission of polygenic traits. Demonstration of cultural transmission would be of great biological and sociological interest. Since there is little evidence for a between families environmental component of variation

or genotype-environment covariance for ability, both of which are generated by cultural transmission of the type envisaged by Eaves (1976b), exploration of this novel aspect of multifactorial inheritance requires a new trait.

Interest in the measurement of personality began when methods for determining an individuals' intelligence were well established. Cattell (1957, 1965a) developed a theory of personality involving sixteen first order factors. This is complex and renders analysis interpretation and prediction difficult. Eysenck's work describing his three personality dimensions - neuroticism, psychoticism and extraversion (Eysenck 1952, 1968; Eysenck and Eysenck, 1969) - has been extensive and provides a more useful model for personality than that of Cattell. Twin data has been used to study the genetical determination of these three major dimensions and simple mechanisms appear to account for variation in neuroticism (e.g. Eaves and Eysenck, 1976), psychoticism (Eaves and Eysenck, 1977) and extraversion (Eaves and Eysenck, 1975). There is no indication of between families environmental variation or genotype-environment covariation. Therefore, interest has turned to the analysis of social attitudes.

Early work showed that the two dimensions of social attitudes - Radicalism - Conservatism and Tough - Tendermindedness - were partly inherited (Eaves and Eysenck, 1974). A priori, we might suppose that these traits are of some cultural and evolutionary significance and that they are suitable candidates for the study of cultural transmission.

In Part B of this thesis a detailed analysis of social attitudes data is given. Analysis of 587 pairs of twins uses conventional methods

for determining genetical and environmental components of variation, but is important for two reasons. The number of twins is now sufficient to overcome the problems of power encountered in the Georgia Twin Study and can 1. allow discrimination between simple environmental and simple genetical models of variation and 2. demonstrate the importance of the joint effects of culture and the mating system. A comparison between the components of variation estimated in Part B and and those from two previous studies is made. The cultural transmission of attitudes based on parental phenotype might be implicated in the production of environmental differences between families, giving rise to genotype-environment covariance. This cannot be studied using data on twins reared together and we again see the need for the collection of data on other relatives. Adoption data are critical for the resolution of cultural transmission and data on 445 individuals who formed part of an adoption study were available. It is in the analyses of these data that there is novelty in the models and analytical techniques used.

The relevance of the model of Eaves (1976b) is considered for these data. However, simultaneous specification of assortative mating and cultural transmission are not possible within the framework of this model. Therefore, new models are formulated using the methods of path analysis. Maximum-likelihood estimation and hypothesis-testing are used in order to estimate parameters of the models and determine which are the most likely explanations of the observed variation. This work is exciting because it combines the use of path models, which have hitherto been used ineffectively in the analysis of human data, with a method of maximum-likelihood estimation which permits the analysis of irregular

pedigrees, (which are normally wasteful of information), in the most efficient manner. It shows how path models can be useful when dealing with effects which are basically phenotypic in origin and demonstrates a trend, likely to prove fruitful, towards the collection of families as they occur naturally in the human population. This will allow a variety of effects to be detected and estimated, without the problem arising that particular sorts of individuals may be atypical of the population as a whole.

However, a large area in human behavioural genetics remains to be discussed. Path models are useful in the analysis of phenotypically based effects (e.g. assortative mating, cultural transmission), but cannot be used to satisfactorily specify non-additive effects such as dominance, epistasis and genotype-environment interaction. A thorough consideration of all the sources of genotype-environmental covariance has not been made. This lead us to an area of great potential importance for the design of future experiments. A series of simulations were performed in order to detect the biases and errors of inference which may occur when certain important effects, such as genotype-environment interaction are mistakenly omitted from models of variation. There is a body of opinion which expresses the view that the heritable component of variation will be over estimated in such cases. The simulations were carried out to show that this is not necessarily the case, although the precise errors made will depend on experimental design, the true causes of variation and the particular mistaken models fitted to the data. In Part C, we outline methods for the exact quantification of the biases for a given situation and show how to estimate the probable errors of inference. The use of these methods in planning future research is considered.

PART A

THE GEORGIA TWIN STUDY

INTRODUCTION

The work reported in the first part of this thesis continues research into an area with a long and controversial history. It concerns the role of biological and environmental factors in producing individual differences in ability. Galton was the first to consider this topic. His studies concerning the importance of inheritance in determining high levels of achievement are summarised in his book, Hereditary Genius (1869), where he showed that among the exceptionally gifted, the number of relatives also outstanding in their achievements is much larger than that expected by chance. Since then a large body of data, collected throughout this century, has suggested that general ability as measured by conventional I.Q. tests is largely genetically determined, at least in Caucasian populations in western Europe and the United States. The evidence provided by these data has been extensively reviewed and discussed recently (e.g. Erlenmeyer-Kimling and Jarvik, 1963; Burt, 1966, 1972; Jarvik and Erlenmeyer-Kimling, 1967; Jensen, 1967, 1969, 1970; Lindzey et al, 1971; Vandenberg, 1971; Jencks, 1972; Anderson, 1974; Broadhurst et al, 1974; Eaves, 1975), and further comments upon it will not be made here.

In the United States, attention became focussed upon the determination of individual differences in intelligence in Black and White populations. Much of this work was summarised in two books edited by Vandenberg (1965, 1968). A large mean difference in ability between the two races was reported (e.g. Shuey, 1966; Jensen, 1969) and work purporting to show that this difference was due to genetical differences caused much argument. Many reviews of the evidence have been published

often reaching very different conclusions (e.g. Klineberg, 1944, 1963; Dreger and Miller, 1960, 1968; Pettigrew, 1964; Shuey, 1966). Much attention was drawn to the debate by Jensen (1969) whose critical review of the research concluded that the mean difference was due mostly to genetical differences between the populations. He showed that I.Q. tests measure a dimension of general ability of great social relevance, that individual differences on this dimension are highly heritable (genetical differences accounting for 80 per cent of the total variance) and that educational programmes have proved generally ineffective in changing the relative status of individuals or groups on this dimension.

The most recent review of the work in this area was made by Loehlin et al (1975) in their book "Race Differences in Intelligence", which provides an excellent account of the background for the present study. They discuss evidence of four main types which may be used to separate genetical and environmental components of variation. Studies of racial admixture, adopted individuals and their parents and half siblings are reported. However data on twins and siblings is the most extensive and will be briefly outlined here.

Jensen (1973) gave a battery of tests of ability to 8000 children and calculated correlations for Black and White sibling pairs. He found similar correlations in Blacks and Whites. In a large study of twin and sibling pairs, Nichols (1970) concluded that the heritability of ability in Blacks was lower than that in Whites. However, the evidence is somewhat ambivalent depending on the age at which the individuals were tested and the particular tests used.

The largest study of Black and White twins is that of Scarr-Salapatek (1971). She had data on aptitude and achievement tests.

However, despite the size of the study, she failed to find a systematically greater similarity of same-sex twins compared with opposite sex twins i.e. there was no demonstrable heritable component. This result is anomalous compared with the results of many previous studies. Thus, of the three studies mentioned so far, two have given internally inconsistent results and the other only tells us that sib correlations are approximately equal in Blacks and Whites.

The present study is a direct follow up of the fourth major study of Black/White differences in ability - that of Vandenberg and Osborne (Vandenberg, 1969; Osborne, Gregor and Miele, 1968).

They administered the same test battery to two sub-samples of twins, one in Louisville, Kentucky and the other in Atlanta, Georgia. Later they independently published papers on the analysis of the combined samples comparing heritabilities of the measures of ability in Blacks and Whites (Osborne and Gregor, 1968; Osborne and Miele, 1969; Vandenberg, 1970). Although both were analysing the same data, they came to contradictory conclusions. Vandenberg found that heritabilities were lower in Blacks, but Osborne found little difference in heritabilities, except that heritabilities were marginally larger in Blacks than Whites for some tests. The discrepancy between the two analyses can, in this case, be explained in terms of the two differing methods of calculating heritability, neither of which is truly the proportion of the total variance asrribable to variation produced by all genetical influences (see Jinks and Fulker, 1970).

Osborne's analyses were based almost entirely on the MZ and DZ twin correlations, whilst Vandenberg used his F statistic, which includes only within families genetical and environmental effects. In both

cases, the choice of statistics and methods of analysis did not make the most efficient use of the available data.

In the following pages, we will analyse the data of the follow-up to their studies, known as the Georgia Twin Study. The methods used will be those of biometrical-genetics, based mainly upon the analysis of variance components. These methods are discussed by Mather and Jinks (1971), Jinks and Fulker (1970), Eaves et al (1977) and others. In this way, we hope to make the most efficient use of the data and to draw more powerful conclusions about the sources of individual differences in ability in Blacks and Whites in the United States.

SECTION 1: A DESCRIPTION OF THE DATA

1 INTRODUCTION

The collection of the data which forms the Georgia Twin Study was begun in the early sixties by Osborne and Vandenberg. Forty five pairs of twins were tested by Osborne in Atlanta, Georgia and Clarke County, Georgia. Two hundred and thirty nine pairs were tested by Vandenberg in Louisville, Kentucky; Jefferson County, Kentucky and Indiana. These 284 pairs of identical and same-sex non-identical twins form the Co-operative Twin Study of Vandenberg (1967) and Osborne (Osborne et al 1968; Osborne and Gregor, 1968).

In 1972, Osborne decided to enlarge the sample size of the Co-operative Twin Study to enable comparisons to be made by race and sex. The earlier study had been criticised by Jensen (personal communication) and Loehlin et al (1975) because the number of Negro pairs (especially dizygotic, Negro pairs) was too small to enable meaningful comparisons to be made. Therefore, in selecting schools for participation in the Extended Study care was taken to ensure that the racial composition of the schools reflected the racial composition of the region. The Extended Sample is composed of 143 same-sex pairs and 47 opposite-sex pairs, tested by Osborne in Atlanta, Georgia; Chatham County, Georgia and Clarke and surrounding counties in Georgia in 1972.

The total sample of 474 pairs of twins from the Co-operative Twin Study and the Extended Sample together form the Georgia Twin Study. The data of the Georgia Twin Study were kindly made available to the

Genetics Department of the University of Birmingham by Dr. R.T. Osborne, Professor of Psychology at the University of Georgia, for biometrical-genetical analysis in the spring of 1974.

2 THE SUBJECTS

2.1 Location of the Subjects

Twins were located through public and private schools in the areas mentioned above. In Kentucky and Indiana, lists of twins were obtained from school boards and permission for the twins to participate in the study was obtained from parents. The twins were paid an honorarium and testing sessions took place during school hours. Vandenberg (1968, 1969) reports that in nearly all cases the twins contacted actually agreed to participate in the study. Co-operation in the schools varied between 85 and 100%. Although co-operation was not 100% in all cases, there was no reason to expect any serious biases introduced by the sampling procedure.

In Georgia location of the twins was by a slightly different procedure. A call for like-sexed twin volunteers in the eighth grade and above was placed through the guidance counsellor in the selected schools. Eighth grade children have an average age of thirteen years. All twin volunteers were accepted as part of the study, including 47 pairs of opposite-sex twins who asked to participate. The consent of the parents and approval of school officials was obtained. Testing took place in two half-day sessions on Saturdays. To encourage both members of the pair to attend the testing sessions and to defray and cost of meals and travelling to school on a Saturday, an honorarium was paid to the twins.

The number of twin pairs not volunteering for the study is unknown, but is believed to be small. Volunteers might not be a random sample

of the population with respect to the traits being measured. However, there was no evidence that this was the case. It was hoped that the volunteers would be a representative sample of twins in schools in the Southern United States.

2.2 Structure of the Sample

The subjects were pairs of monozygotic and dizygotic twins. Forty seven pairs of dizygotic twins were opposite-sexed. These twins had volunteered for the study when the call for like-sexed twins went out. It was decided to test them also because turning away some volunteers might discourage others from participating. However, since their inclusion was accidental, it was decided to exclude these pairs from the analysis in case they were a biased sample of opposite-sexed pairs. Analysis of data on the same-sexed pairs was performed and is reported here.

The twins were further classified as male or female and into two racial groups, i.e. whites and blacks. The ages of the twins lie in the range from twelve years to twenty years. The mean age is 15.33 years with a standard deviation of 1.56 years. This age range is an additional factor in ensuring that twins could not have been included in 1963 and 1972. Testing took place in different schools on the two occasions, but a pair might possibly have been included twice by transferring from one school to another. However, all twins tested in 1963 would have completed their schooling by the time of the second testing in 1972.

The distribution of the total sample by age, race and sex is shown in Table A1. Sixty-three pairs of twins, for whom there were missing data for one or both members of the pair, are excluded from the analysis. Thus, a sample of 364 pairs of twins remains. Their distribution by zygosity, race and sex is shown in Table A2.

TABLE A1: DISTRIBUTION OF THE TWINS BY RACE, SEX AND AGE

Age	White			Negro			Total			Twin Pairs	
	Male	Female	Total	Male	Female	Total	Male	Female	Total	No.	%
12		2	2		2	2		4	4	2	.4
13	30	46	76	14	30	44	44	76	120	60	14.1
14	46	52	98	24	36	60	70	88	158	79	18.5
15	54	76	130	18	34	52	72	110	182	91	21.3
16	54	74	128	16	32	48	70	106	176	88	20.6
17	56	52	108	6	24	30	62	76	138	69	16.2
18	22	34	56	2	8	10	24	42	66	33	7.7
19	6	2	8				6	2	8	4	.9
20	2		2				2		2	1	.2
Means	15.59	15.41	15.49	14.78	15.01	14.93	15.41	15.28	15.33		
S.D's	1.61	1.56	1.58	1.29	1.50	1.44	1.58	1.55	1.56	427	
No.Twin Pairs	135	169	304	40	83	123	175	252	427		

TABLE A2: DISTRIBUTION OF THE TWINS BY ZYGOSITY, RACE AND SEX

	Monozygotic Twins				Dizygotic Twins			
	Males	Females	Total		Males	Females	Total	Grand Totals
Caucasians	66	75	141	Caucasians	42	68	110	251
Negroes	22	47	69	Negroes	12	32	44	113
Totals	88	122	210	Totals	54	100	154	364

3 DIAGNOSIS OF ZYGOSITY

All twins studied in 1963 were blood-typed by the Minneapolis War Memorial Blood Bank. The following factors were tested: A, B, O, M, N, S, s, P_1, Rho, rh'', rh', Miltenberger, Vermeyst, Lewis, Lutheran, Duffy, Kidd, Sutter, Martin, Kell, Cellano and occasionally some others. All pairs discordant for one or more blood types must be dizygotic. However, a small proportion of dizygotic twins may be concordant for all blood types tested just by chance and therefore, be misclassified as monozygotic. This will increase the within pairs variance of monozygotic twins and so the effect of within families environmental influences will be over estimated. Thus, misclassification produces a bias which reduces the proportion of variance ascribable to genetic factors.

Misdiagnosis of zygosity is a possible explanation for the significant excess of monozygotic twins compared with the numbers expected from the proportion of monozygotic twins-births in the population given by Standskov and Edlen (1946) (see section 1.5.2.). However, Smith and Penrose (1955) give tables for the most frequently tested blood groups which enable the probability of misclassification to be calculated. Martin and Martin (1975) found probabilities for concordant pairs being dizygotic varying between 0.002 and 0.035. Assuming a fairly high estimate of three per cent misclassification, then in Vandenberg's sample we would expect only seven pairs of twins concordant on all blood markers to be dizygotic. Since there were twenty pairs of monozygotic twins more than expected, it seems that although misclassification may contribute to the excess of monozygotic twins, it is not a sufficient factor in itself to explain the excess. The proportion of

monozygotic:dizygotic twins and the excess of monozygotic twins is further discussed in Section 1.5.2.

Vandenberg classified the 239 pairs of twins studied by him as monozygotic or dizygotic using the results of serological tests alone. Possible misdiagnosis could have led to as many as seven pairs being misclassified. This is a small proportion of the total sample and will produce only a small bias. However, such misclassification will lead to the under-estimation of the proportion of the variance due to genetical causes.

Osborne had anthropometric as well as serological data for the sub-sample of forty five pairs of twins studied by him in 1963. The anthropometric measures were: face length, head length, head breadth, head circumference, height and weight. Three other measures were computed: Cephalic index, Kaup's Index (body weight in grams/(height (in cm))^2) and Rohrer's Index of body structure (body weight x 100/(height (in cm))^3). Data on colour-blindness and handedness were also available. Twins were automatically classified as dizygotic if they were discordant for blood-type or colour blindness. Concordant twins were then classified as similar (MZ) or dissimilar (DZ) using the nine physical measures described above. Six pairs of twins concordant for all blood-types were classified as dissimilar and called dizygotic on the basis of probabilities obtained from the tables of Verschuer (1932) and Dahlberg (1926). It was hoped that this procedure would pick out dizygotic twins who would otherwise have been classified as monozygotic because they were concordant on all blood-types by chance. Reclassification of six pairs out of forty five on the basis of physical measures implies thirteen per cent misclassification by blood-typing, which is higher than in any

previous study. When we consider the absolute numbers involved, we find that we would have expected to reclassify one or two pairs on the basis of physical measures and we have, in fact, reclassified six pairs. Thus, even if we have mistakenly classified some monozygotic twins as dizygotic by using the physical measures, the absolute number of pairs involved is small and unlikely to produce a large bias. This procedure would produce misdiagnosis of the type where true monozygotic twins are mistakenly called dizygotic because they differ more in their physical characteristics than expected for monozygotic twins. They will not necessarily differ more than expected for the abilities being studied here and, therefore, the within pairs variance of dizygotic twins will be reduced. This means that the within families genetical effect and the proportion of variance due to genetical influences will be underestimated. Thus, the possible sources of misclassification in Vandenberg's sub-sample and Osborne's sub-sample studied in 1963 both lead to underestimation of the genetical components of variance.

The twins studied by Osborne in 1972 were not blood-typed. Physical measures and characteristics were used in conjunction with two questionnaires to determine the zygosity of these twins. One questionnaire was that developed by Nichols and used in the National Merit Twin Study (Nichols, 1965; Nichols and Bilbro, 1966); the other was a modification of that used by Schoenfeldt for Project Talent. (Schoenfeldt, 1968).

On the basis of this data, two computer programs were used to classify the pairs as monozygotic or dizygotic. The Automatic Interaction Detector (AID) Program is explained in detail by Schoenfeldt (1968). The other program was the Discriminant Analysis Program (BMD07M) from

the Package of Biomedical Computer Programs (Dixon, 1973). This utilised the nine physical measurements described for the twins studied by Osborne in 1963. The 1963 group were used as the criterion group for determining the accuracy of diagnosis using the discriminant scores of the 1972 group. The results of the two programs were compared. Where these agreed, the twin pair was classified accordingly. Of the 143 pairs, 61 pairs were classified as monozygotic and 35 as dizygotic. This total of 96 pairs had been classified the same by both programs.

For the 47 pairs of twins remaining, a second discriminant analysis was run using the 96 pairs of twins already diagnosed as the criterion group. Seventeen variables were used in this analysis. They were the nine measures from the first analysis plus nose length, eye colour, hair colour, other hair differences, colour blindness (two variables), mistaken identity variables from the Project Talent Questionnaire and handedness.

Three judges classified the 47 pairs of twins as monozygotic or dizygotic using front and profile photographs, statements of likenesses and differences made by the twins and the twins self-report of zygosity. Thirty six pairs of twins were diagnosed as identical by the second discriminant analysis and by the judges, these were classified as such (20 MZ pairs and 16 DZ pairs).

The complete files of the eleven remaining pairs (except the psychological test results) were examined by the principal investigator and he made the final classification of these pairs. His comments are presented in Appendix A.

The results of the AID program agreed with the final result in 81.2 per cent of cases. For the nine variable and seventeen variable discriminant analysis programs the diagnoses agreed with the final classification in 85.3 and 76.6 per cent of cases respectively. Assuming the final classification to be the correct one, the reliability of all these methods is low. However, in the 47 cases where judges were used, the judges consensus always agreed with the final classification. Out of the eleven cases, where no decision could be made and the principal investigator decided the issue, ten pairs turned out to be dizygotic. These comments on the zygosity determination are especially interesting if we consider several studies of zygosity diagnosis. Cederlof et al (1961) recorded an accuracy of 98.6 per cent with 145 pairs of twins using two mailed questionnaires about physical similarity when growing up and confusion in childhood. Unfortunately they could make no diagnosis for 55 (or 27.5 per cent) of the original 200 pairs either because of failure to reply or because diagnosis could not be made from the responses given. More recent work by Kasriel and Eaves (1976) showed an accuracy of 96.1 per cent using only two questions on physical similarity and confusion in childhood. The "true" zygosity of the twins was determined by blood-typing 178 pairs of twins for fifteen different systems. Of 94 pairs of twins diagnosed as monozygotic by blood-typing, 92 pairs agreed that they were alike and had been confused in childhood. If pairs agreed that they were alike and were confused in childhood, Kasriel and Eaves classified them as monozygotic. If they said they were not alike and not confused, or if they failed to agree on these two questions, they were classified as dizygotic. Using these criteria only 7 pairs out of the 178 pairs were misdiagnosed (i.e. 3.9 per cent).

In the Georgia Twin Study, the judges and the principal investigator used responses to questions on similarity and confusion in childhood to help reach their decision. Kasriel and Eaves point out that disagreement between the twins as to their zygosity and any other doubts about zygosity are usually a good indication that the pair in question is dizygotic. In this study, we notice that of the final eleven pairs for whom there was doubt about zygosity, ten pairs were finally classified as dizygotic.

It is hoped that the lengthy processes involved in reaching a decision about the zygosity of these 143 twin pairs has reduced errors in diagnosis to a very low level. Errors could have occurred in either direction, i.e. some monozygotic twins may have been classified as dizygotic and some dizygotic twins may have been classified as monozygotic. The effect of these errors would be to increase the within pairs variance of the monozygotic twins, leading to overestimates of the within families environmental effect, and to decrease the within pairs variance of the dizygotic twins, leading to underestimates of the within families genetical effect. Both these effects will reduce estimates of the 'proportion' of variance due to genetical influences.

Zygosity was determined differently in the three sub-samples studied, leading to misclassification of dizygotic pairs as monozygotic in Vandenbergs sub-sample, misclassification of monozygotic pairs as dizygotic in Osborne's 1963 sub-sample and errors in both directions in Osborne's 1972 sub-sample. However, in each case the bias introduced by the misdiagnosis of zygosity is predictable and leads to the underestimation of the proportion of the variance due to genetical causes.

4 THE TESTS

All the tests selected for use in the Georgia Twin Study were easy to administer by pencil and paper methods. The twins were tested in groups and testing usually took place in the subjects own school in two half day sessions.

Thirteen tests of ability were taken by all the twins participating in the study. These tests were chosen to represent the unique, primary mental abilities identified independently by Thurstone, Cattell and Guilford. Tests thought to include items requiring specific learned achievements or tests of short term memory were not considered for use in the Georgia Twin Project. These tests measure abilities called Level I abilities by Jensen. Level I abilities are similar in nature to Cattell's crystallised general ability (g_c) which is a general factor found "largely in a type of abilities learned at school, representing the effect of past application of fluid intelligence, and amount and intensity of schooling". (Cattell, 1965a). Jensen's Level I abilities and Cattell's crystallised general ability have in common an emphasis on learned achievements.

All the tests used in the Georgia Twin Project should be measuring the Level II abilities of Jensen. "Level II ability is characterised by transformation and manipulation of the stimulus prior to making the response. It is the set of mechanisms which makes generalisation beyond primary stimulus generalisation possible. Semantic generalisation and concept formation depend upon Level II ability; encoding and decoding of stimuli in terms of past experience, relating new learning to old learning, transfer in terms of concepts and principles, are all examples

of Level II. Spearman's characterisation of g as the 'eduction of relations and correlates corresponds to Level II". (Jensen, 1973a). It is also similar to Cattell's fluid intelligence (g_f) which is "that form of intelligence which is largely innate and adapts itself to all kinds of material, regardless of previous experience with it". (Cattell, 1965a). However, it seems unlikely that all the tests chosen for use in the Georgia Twin Study are entirely free of Level I abilities, especially tests of verbal ability and arithmetic ability.

Six of the tests used in the Georgia Twin Study were selected from the Educational Testing Service (ETS) kit of reference tests, but the others are less well-known. All were originally selected by Vandenberg for use in the Louisville Twin Study (Vandenberg, Stafford and Brown, 1965). The tests measuring each primary mental ability are shown in Table A3, together with other information about the primary mental abilities involved. A brief description of the thirteen tests used in the Study, reference to their source and the abbreviation for each used in the text are given in Appendix B.

There are data only on these thirteen tests for the sub-sample of twins studied by Vandenberg in 1963. The forty-five pairs of twins tested by Osborne in 1963 took nine additional tests. These are shown in Table A4 together with their sources. The sub-sample tested in 1972 took the thirteen main tests and also four others as follows:

1. Test of "g" culture fair, A
2. Test of "g" culture fair, B

(1 and 2 prepared by Cattell and Cattell)

3. Primary Mental Abilities Test (PMA) prepared by Thurstone
4. The Junior-Senior High School Personality Questionnaire (J.S.HSPQ), prepared by the Institute for Personality and Ability Testing.

TABLE A3: EXPECTED FACTOR STRUCTURE OF TESTS USED IN THE GEORGIA TWIN STUDY

	Investigators identifying the Factors	General types of test	Specific test used in the Georgia Twin Study
FACTOR I VERBAL ABILITY	Thurstone (1938) Burt (1940) Garrett (1946) Cattell (1957) Guilford (1967)	Vocabulary Reading Comprehension Grammar etc.	Wide range vocabulary Self-judging vocabulary Spelling Calendar test
FACTOR II NUMERICAL ABILITY	Thurstone (1938) Garrett (1946) Cattell (1957) Guilford (1967)	Addition Subtraction Multiplication Division etc.	Simple arithmetic
FACTOR III SPATIAL ABILITY	Thurstone (1938) Fruchter (1948) Macfarlane Smith (1954) Guilford and Lacey (1947)	Formboard Punched Holes Cubes etc.	Newcastle Spatial Cubes comparison Surface development Paper folding Formboard Object aperture test
FACTOR IV PERCEPTUAL SPEED	Thurstone (1938) Cattell (1957) Guilford (1967)	Speed in finding figures and making comparisons	Mazes Identical pictures

The socio-economic status of the parents of these twins was determined also. A summary of all the tests used in the Georgia Twin Project is given in Table A4.

TABLE A4: THE TESTS USED IN THE GEORGIA TWIN STUDY

TESTS TAKEN BY ALL TWINS		TESTS TAKEN BY TWINS OSBORNE IN 1963
TEST	SOURCE	TEST
Arithmetic	Mukherjee (1963)	Card Rotation
Vocabulary	French et al (1963)	Cancellation
Heim	Heim et al (1965)	Whiteman Test of Social Perception
Spelling	Allen et al (1946)	Logical Reasoning
Calendar	Remondino (1962)	Ship Destination
Identical pictures	French et al (1963)	Mooney Faces
Cubes	French et al (1963)	Draw a Man
Surface development	French et al (1963)	Inference I
Form Board	French et al (1963)	How well do you know yourself?
Paper folding	French et al (1963)	
Object aperture	DuBois and Gleser(1948)	
Spatial ability	Macfarlane Smith and Lawes (1960)	
Mazes	Mackinnon and Henle (1948)	

TESTED BY	TESTS TAKEN BY TWINS TESTED BY OSBORNE IN 1972	
SOURCE	TEST	SOURCE
French et al (1963)	Test of 'g' Culture Fair A	Cattell(1959)
Van Der Ven (1965)	Test of 'g' Culture Fair B	Cattell(1959)
Whiteman (1954)	Primary Mental Abilities	Thurstone (1938)
Hertzka and Guilford (1965)	Junior-Senior High School Personality Questionnaire	Cattell(1967)
Christensen and Guilford (1955)		
Mooney (1957)		
French et al (1963)		
Jenkins et al(1959)		

5 ASCERTAINMENT OF THE SAMPLE

5.1 Sources of Bias

In studies such as this, one hopes to make inferences about a population by analysing data from a sample of the population. Therefore, ideally, the sample should be drawn at random to avoid introduction of biases. When human subjects are used, sampling is rarely random for a number of reasons.

Volunteering Behaviour

When research is carried out on human populations, all subjects are volunteers. This immediately introduces the problem that volunteers may not be typical with respect to the trait under study. For example, it is well known that females tend to volunteer more frequently than males.

It is also known that females score more highly on average on neuroticism scales than males (Eysenck and Eysenck, 1969). Therefore, in any sample, the number of subjects with high neuroticism scores will be over-represented. In this case the problem is easily overcome. The two sexes are recognised as separate classes and tested for mean differences and differences in total variance. Similarly racial differences and differences in socio-economic status may be recognised and tested. However, less obvious classifications of the data may exist. If there are mean differences between the classes and some classes tend to volunteer more often than others, then unsuspected biases may be

introduced into the data. The problem is more serious where psychological traits are being studied, and volunteering behaviour may be partly determined by the trait being studied.

The Representativeness of Twins

In the Georgia Twin Study problems could be introduced by the constraint, inherent in the experimental design, that all subjects must be twins. If twins are not typical of the population from which they are drawn with respect to the trait under study, then conclusions based on the sample may not be legitimately generalised to the population as a whole.

Several workers have criticised the "Twin Method", because twins are atypical. Evidence that twins are atypical for the type of abilities studied here has come from a number of sources. Various workers (Mehrotra and Maxwell, 1949; Sandon, 1957; Drillien, 1961; Husén, 1969; Record, McKeown and Edwards, 1970) have shown that twins have a mean for tests of ability several points lower than that of singletons. Mittler (1971) reviews the extensive evidence for the lower mean ability of twins. Several workers (e.g. Sandon, 1957) have shown that there are relatively more individuals with low ability among twins than among singletons which suggested that the operation of pre-natal and natal factors may cause cerebral injuries more often among twins than among singletons. Price (1950) has fully reviewed the literature on the possible sources of bias introduced by pre-natal and natal factors in twins. Husén (1969) in his study of 2200 singletons and 1000 individual twins has shown that the lower mean score of twins is due

to a downward displacement of the whole ability distribution among twins, which makes the birth injury hypothesis less tenable. Husén (1959) showed that twins tend to display a verbal retardation and suggested that psychological factors gave rise to the lower mean ability, since twins seemed to communicate more among themselves and less with the outside world than singletons. He predicted that monozygotic twins would be more retarded than dizygotic twins. However, Husén showed in his 1969 paper that this was not in fact the case and that identical twins show somewhat higher means.

Several workers (e.g. Willerman and Churchill, 1967) have pointed out the correlation between low birth weight and ability and the lower birth weight of twins and suggested that low birthweight and ability are causally connected.

Record, McKeown and Edwards (1969a, b; 1970), however, have shown in a large study that the low ability scores of twins are not explained by differences from single births in their distribution by maternal age and birth order or by birth weight and duration of gestation. They are also not explained by any increased risks in monozygotic twins or with delivery of the second twin. They suggest that the explanation for the difference in ability between twins and singletons must be sought in the post-natal environment.

They have also compared verbal reasoning scores of twins raised as singletons due to the death of the co-twin before or soon after birth with the scores of twins reared in pairs. In this case the prenatal environments of twins raised as singletons and twins raised in pairs were similar. Only the post-natal experiences of the two groups differed. The twins raised as singletons had higher scores than the

twins raised as pairs, which were only slightly lower than those of singletons. This suggests that the cause of the lower mean intelligence scores of twins must be sought in differences in post-natal environment between twins and singletons. Whatever the cause of the lower mean scores of twins, it seems that twins are not typical of the population as a whole with respect to environmental factors causing variation in ability. Therefore, great caution must be exercised in generalising the results of twin studies to the population as a whole.

Other sources of Bias in the Georgia Twin Project

Given that the "twin method" is to be used, there are many other possible sources of bias. The first is that the schools chosen for study may not be random with respect to the abilities of interest. However, a number of both private and public schools over a wide area were used in this study and bias from this source, if present, should be small. Secondly individuals may not be chosen randomly within schools. Vandenberg tried to eliminate this source of bias by testing all individuals in the schools studied. However, cooperation was not a hundred per cent in all cases and twins opting out might not be random with respect to the traits of interest. However, cooperation was high varying from 85 to 100 per cent (Vandenberg et al, 1968). Therefore, any biases are likely to be small.

Osborne did not attempt to contact all twins through lists obtained from school boards. He put out a call for volunteers in the schools selected for study. The sample obtained in this case will only be random if volunteers are a random selection of twins, as discussed above.

Although there was no evidence on this point, Osborne believed the number of non-volunteers to be small. Therefore, any biases are likely to be small.

Possible deviations from random sampling have been discussed here. At the time of testing there was no reason to suspect any serious biases being introduced by the sampling procedures. However, tests of the randomness of sampling will be made.

5.2 Tests of Sampling

The observed numbers in the sub-groups of the data are given in Table A2. Table A5 shows the observed numbers in the sub-groups further divided into three sub-samples. There are several simple tests of sampling suggested by the structure of these data.

The Racial Composition of the Sample

In Georgia, the Black population forms 26 per cent of the total population and in Kentucky the Black population forms only 12 per cent of the total population (Osborne, personal communication). A simple chisquare test of the observed numbers of Whites:Blacks against the expected numbers, calculated from the population proportions was carried out for each sub-sample separately. In the sub-sample collected by Vandenberg in 1963, the observed numbers of each race do not differ significantly from those expected, given that 12 per cent of the population in Kentucky is Black ($X_1^2 = 1.28$). However, in both bodies of data collected by Osborne, there are more Blacks than expected on the

TABLE A5: COMPOSITION BY RACE, SEX AND ZYGOSITY OF THE THREE SUB-SAMPLES OF THE DATA

A. OSBORNE '63

	CAUCASIANS				NEGROES		
	MZ	DZ	Total		MZ	DZ	Total
Males	4	7	11	Males	3	2	5
Females	9	5	14	Females	10	5	15
Total	13	12	25	Total	13	7	20

B. OSBORNE '72

	CAUCASIANS				NEGROES		
	MZ	DZ	Total		MZ	DZ	Total
Males	17	13	30	Males	15	9	24
Females	17	16	33	Females	33	23	56
Total	34	29	63	Total	48	32	80

C. VANDENBERG '63

	CAUCASIANS				NEGROES		
	MZ	DZ	Total		MZ	DZ	Total
Males	63	31	94	Males	8	3	11
Females	61	61	122	Females	7	5	12
Total	124	92	216	Total	15	8	23

basis of random sampling given that 26 per cent of the population in Georgia is Black ($\chi_1^2 = 7.96$ for the 1963 sample and $\chi_1^2 = 66.64$ for the 1972 sample). Therefore, sampling in Georgia was not random as far as race is concerned.

In the Southern United States, the children in any school belong predominantly to one racial group. The high proportion of Blacks in Osborne's samples was probably obtained by choosing a large number of schools where Blacks predominate. Since we are interested in a comparison of the two races, the excess of Negroes in the sample, produced by selection of schools, will not bias the analysis in any way.

The Sex Ratio in the Sample

The expected ratio of males to females is 1:1, assuming that viability and survival rate are the same in the two sexes. The excess of females in the total sample is highly significant ($\chi_1^2 = 17.58$), but the excess is only significant in one sub-sample - that tested by Osborne in 1972. Osborne (personal communication) has suggested a reason for this. Many boys in the age group being studied have Saturday jobs and would be unable to participate in the study. Far fewer girls have jobs and so more girls would be free to attend the testing sessions. This is probably a factor contributing to the excess of females in Georgia, but in Kentucky and Indiana testing took place during school hours. The well known greater volunteering frequency of females is probably the main explanation of the excess of females in the sample (see section 1.5.1). The excess of females in the sample could bias the analysis if the sexes differ in the abilities measured here

and if the sexes are pooled. Whether males and females do differ in these abilities will be tested during the preliminary analysis.

Relative Frequencies of Monozygotic and Dizygotic Twins

The expected proportion of monozygotic:dizygotic twins may be found using Weinberg's rule (Weinberg, 1901). All monozygotic twin pairs are like-sexed, but assuming the sex ratio is 1:1, half the dizygotic twin pairs will be same-sexed and half opposite-sexed. Therefore, knowing only the numbers of same-sexed and opposite-sexed twin births, the proportion of monozygotic:dizygotic twin pairs may be found as follows (Bulmer, 1970):

$$\text{Proportion of MZ twin pairs} = \frac{L-U}{N}$$

$$\text{Proportion of DZ twin pairs} = \frac{2U}{N}$$

where L = number of like-sexed pairs

U = number of unlike-sexed pairs

N = number of maternities

Using Weinberg's rule in their study of 31 million multiple and single births, Standskov and Edlen (1946) found that 33.46 per cent of all twin pairs were monozygotic and that 34.17 per cent of White twins and 28.89 per cent of Black twins were monozygotic.

There are no opposite-sexed pairs included in the analysis of this study. Assuming that the number of like-sexed twins equals the number of opposite-sexed twins, then those 37.92 per cent of all White Twins who are opposite-sexed, have not been ascertained at all (per cent

unascertained = (100-34.17)/2). Similarly 35.56 per cent of all Black twins, who are opposite-sexed have not been ascertained. Therefore, we expect 50.94 per cent of White like-sexed twins to be monozygotic (per cent MZ = 34.17/(34.17 +(100-34.17)/2) and, similarly, 44.83 per cent of Black like-sexed twins to be monozygotic. In this study, 56.18 per cent of White twins were monozygotic and 61.06 per cent of Black twins were monozygotic. These percentages are higher than expected for both races. However, comparing the observed numbers of monozygotic and dizygotic twins in this study with the numbers expected on the basis of the proportions of the two zygosities among like-sexed twins, given above, we find that the number of monozygotic twin pairs is not significantly greater than that expected in the Whites ($\chi_1^2 = 2.76$) but there are significantly more monozygotic twins than expected among the Blacks ($\chi_1^2 = 12.04$).

The misclassification of some dizygotic twin pairs as monozygotic could have contributed to the excess of monozygotic twin pairs, which was significant in the Black sample. However, twins were not classified as monozygotic unless they were concordant for a wide range of blood groups and physical characteristics (see Section 1.3). It seems unlikely that misclassification could have produced such large deviations from the expected proportions of the two types of twin pair. Over-ascertainment of monozygotic twin pairs has been found in a number of studies (Nichols, 1965; Schoenfeldt, 1968; Vandenberg et al, 1968) and misclassification of dizygotic twins cannot explain this fully. Another possible reason for the excess of monozygotic twin pairs is that ascertainment procedures may miss the most discordant twin pairs.

A large difference in ability, might mean that co-twins attended

different schools or were in different grades. Among the older twins it could mean that one remained at school, while the other was already at work. Large differences in personality might reduce the probability of both twins volunteering for the study. All these factors may mean that the most discordant pairs are lost to the study. Since the most discordant pairs will be dizygotic, if ability and personality have a genetic basis, then more dizygotic twins will be lost to the study than monozygotic. If this is part of the cause of the excess of monozygotic twins in the study, then the biases produced are predictable. The within pairs variance of dizygotic twins will be reduced and, therefore, the within families genetical component will be underestimated and the proportion of the variance due to genetical influences will be underestimated.

We have shown that the sample is not random with respect to race, sex or zygosity. However, the possible biases this may produce will be explored in the course of the analysis.

5.3 Comparison of the Three Sub-Samples

The total sample to be analysed here was obtained by pooling data from the three sub-samples which were drawn from different locations at different times as described above in the introduction. Such sub-samples may only be legitimately pooled in this way only if they are homogeneous and may be regarded as belonging to the same population.

The means and variances for the three sub-samples on each test by race and sex are presented in Tables A6a to A6d. A simple one-way analysis of variance between the within sub-samples (Snedecor and

TABLE A6a: MEANS AND VARIANCES OF THE SUB-SAMPLES -

TEST		Osborne 1963	Osborne 1972
Arithmetic	Mean	81.1364	62.3500
	Variance	1086.6948	658.0280
Vocabulary	Mean	6.5000	5.2667
	Variance	12.9286	7.9955
Heim	Mean	50.0909	47.1167
	Variance	326.8485	337.2573
Spelling	Mean	37.3182	29.0167
	Variance	87.6558	162.5929
Calendar	Mean	14.8182	12.8667
	Variance	61.9654	49.6090
Identical Pictures	Mean	64.7727	67.1500
	Variance	144.6602	297.9602
Cubes	Mean	11.9545	7.9667
	Variance	102.6169	80.7785
Surface Development	Mean	34.2727	23.5667
	Variance	207.8268	191.6057
Form Board	Mean	16.5000	14.8833
	Variance	60.6429	55.7997
Paper Folding	Mean	9.8636	9.2500
	Variance	27.1710	21.7161
Object Aperture	Mean	9.2273	6.5667
	Variance	83.0411	38.4870
Spatial Ability	Mean	67.4545	61.7333
	Variance	452.1645	335.3853

- WHITE MALES

Vandenberg 1972	F	χ^2	Significance level
61.5134	25.84		***
550.7350		5.32	NS
4.3245	11.33		***
11.9637		3.61	NS
32.2097	32.70		***
353.0639		0.09	NS
28.8486	16.41		***
231.7487		8.64	*
12.2567	11.49		***
45.1273		1.10	NS
48.5414	112.42		***
118.8830		21.25	***
9.3967	5.11		**
72.9182		1.28	NS
24.7074	17.05		***
162.5503		1.04	NS
11.8289	17.37		***
47.4974		1.01	NS
7.7540	12.99		***
21.7779		0.50	NS
6.1290	5.04		*
41.6049		6.02	NS
54.8598	34.61		***
347.0784		0.79	NS

TABLE A6b: MEANS AND VARIANCES OF THE SUB-SAMPLES - WHITE FEMALES

TEST		Osborne 1963	Osborne 1972	Vandenberg 1963
Arithmetic	Mean	73.1786	69.8788	63.2273
	Variance	1091.4855	749.0620	483.7531
Vocabulary	Mean	4.7500	4.9545	4.6148
	Variance	14.3426	10.9364	11.4559
Heim	Mean	49.5000	45.5758	33.4444
	Variance	216.6296	227.6634	262.2066
Spelling	Mean	35.7143	34.0758	35.7500
	Variance	120.2041	118.1003	159.9662
Calendar	Mean	13.9643	12.5303	12.0083
	Variance	38.7765	43.3606	35.0916
Identical Pictures	Mean	65.3571	66.1364	49.6513
	Variance	142.4603	282.0888	120.9369
Cubes	Mean	10.5000	6.6212	6.6750
	Variance	78.5556	67.7774	66.2036
Surface Development	Mean	26.9643	26.9848	21.8484
	Variance	124.1098	115.5228	95.8411
Form Board	Mean	15.1071	12.5758	9.9502
	Variance	37.8029	44.8942	32.0058
Paper Folding	Mean	8.7857	7.9545	7.5925
	Variance	23.8042	21.8287	14.3581
Object Aperture	Mean	4.0000	3.0455	3.1694
	Variance	37.2593	28.3517	21.3363
Spatial Ability	Mean	59.8929	53.5303	48.0693
	Variance	262.0251	344.9298	283.5082

F	X^2	Significance level
3.56		•
	12.67	••
0.26		NS
	0.78	NS
24.41		•••
	0.80	NS
0.50		NS
	2.51	NS
1.37		NS
	1.21	NS
57.63		•••
	21.01	•••
2.80		NS
	0.37	NS
8.76		•••
	1.53	NS
12.83		•••
	3.28	NS
1.17		NS
	7.03	•
0.41		NS
	5.55	NS
7.50		•••
	1.19	NS

TABLE A6c: MEANS AND VARIANCES OF THE SUB-SAMPLES -

Test		Osborne 1963	Osborne 1972
Arithmetic	Mean	39.9000	38.6667
	Variance	723.6556	870.0993
Vocabulary	Mean	4.2000	3.6875
	Variance	5.0667	8.3896
Heim	Mean	39.4000	26.3750
	Variance	330.9333	317.4734
Spelling	Mean	24.9000	20.8958
	Variance	204.7667	221.9251
Calendar	Mean	9.6000	6.5625
	Variance	33.8222	26.1662
Identical Pictures	Mean	47.7000	50.5000
	Variance	199.1222	305.7447
Cubes	Mean	-1.1000	3.6458
	Variance	39.2111	58.7868
Surface Development	Mean	11.2000	17.5625
	Variance	9.9556	63.6981
Form Board	Mean	7.4000	6.7917
	Variance	27.3778	14.7642
Paper Folding	Mean	2.9000	4.3958
	Variance	11.8778	15.1804
Object Aperture	Mean	1.7000	2.9375
	Variance	13.7889	29.2088
Spatial Ability	Mean	42.1000	41.5417
	Variance	349.2111	291.0621

- BLACK MALES

Vandenberg 1972	F	χ^2	Significance level
40.6364	0.04		NS
509.4805		1.88	NS
2.8636	1.13		NS
4.0281		3.80	NS
16.5238	6.36		**
198.1619		1.52	NS
18.4091	0.68		NS
213.2056		1.83	NS
7.2381	1.24		NS
40.4905		1.41	NS
39.2381	3.45		*
213.9905		1.22	NS
4.2727	2.05		NS
47.7316		0.72	NS
16.7727	2.38		NS
112.8506		12.51	**
8.9524	1.59		NS
34.7476		5.78	NS
4.6818	0.74		NS
18.8939		0.71	NS
3.5000	0.43		NS
24.1667		1.82	NS
41.7143	0.00		NS
387.5143		0.67	NS

TABLE A6d: MEANS AND VARIANCES OF THE SUB-SAMPLES - BLACK FEMALES

Test		Osborne 1963	Osborne 1972	Vandenberg 1972
Arithmetic	Mean	40.1667	50.1161	51.7917
	Variance	582.9713	819.1306	463.9982
Vocabulary	Mean	2.1667	3.0893	3.9583
	Variance	9.4540	9.2712	4.9112
Heim	Mean	27.5333	26.4107	22.0000
	Variance	373.2230	257.2893	177.3913
Spelling	Mean	27.9000	27.3571	32.2500
	Variance	298.7140	208.1956	179.9348
Calendar	Mean	8.7333	6.8393	6.3333
	Variance	57.5126	36.9109	32.7536
Identical Pictures	Mean	53.1333	55.3571	43.3500
	Variance	177.0851	373.2767	209.5026
Cubes	Mean	-0.6333	0.4018	4.0833
	Variance	62.1713	38.4587	48.1667
Surface Development	Mean	16.4333	17.0625	17.1250
	Variance	35.2195	49.6627	39.9402
Form Board	Mean	7.1000	6.0625	6.1667
	Variance	20.0241	13.7528	8.3188
Paper Folding	Mean	4.5000	3.0625	4.1667
	Variance	17.4310	14.7438	18.7536
Object Aperture	Mean	0.5333	0.5804	3.1250
	Variance	12.0506	12.6061	17.1576
Spatial Ability	Mean	36.3000	34.0893	34.8750
	Variance	195.0448	186.5145	193.2446

F	χ^2	Significance level
1.81		NS
	3.44	NS
2.50		NS
	3.42	NS
0.89		NS
	3.54	NS
1.08		NS
	2.08	NS
1.27		NS
	2.91	NS
4.60		*
	6.97	*
3.83		*
	2.97	NS
0.11		NS
	1.48	NS
0.91		NS
	4.69	NS
1.97		NS
	0.75	NS
4.25		*
	1.08	NS
0.31		NS
	0.03	NS

Cochran, 1967) was performed for each test for the four race-sex groups separately, so that any significant differences between the means of the three sub-samples could be detected.

The results of all the analyses are summarised in Tables A6a to A6d. The values of the F ratios for testing the assumption that there are no mean differences between the three sub-samples are shown for each in Tables A6a to A6d together with their corresponding significance levels.

Differences between the means of the three sub-samples were highly significant for White males. The differences between the means of the White females were highly significant for seven of the tests. However, there was little evidence for significant differences between the means of the three sub-samples in either Black males or Black females. Differences between the means of the sub-samples were significant for two tests only in each sex.

The mean differences between the White sub-samples could be genuine population differences or could be explained if sampling were not from the complete distribution in one or two of the sub-samples. Since there was some evidence for non-random sampling in these data (see Section 1.5.2), it was decided to test the variances of the sub-samples for homogeneity in case there was any restriction in the range of variation present. The total variances of the three sub-samples were tested for homogeneity using Bartlett's Test. The values of chisquare for testing the assumption that the variances of the three sub-samples are homogeneous are shown in Tables A6a to A6d together with their corresponding significance levels.

There is little evidence for heterogeneity of total variances in either Whites or Blacks. Therefore, there is no evidence for restriction

in the range of variation in any of the sub-samples. The difference in mean between the three White sub-samples are probably genuine population differences between the sub-samples. This suggests that the Blacks form a homogeneous racial group but that the Whites are heterogeneous and the sub-samples belong to distinct and different populations.

When the data are analysed as a whole, the analysis of the Black group is not biased in any way by pooling the three sub-samples of the data. However, in the Whites, the between sub-samples differences will contribute to the between families differences which will be biased in an upward direction as a result. The between families differences will be inflated and the importance of both between families genetical influences and between families environmental differences will be over-estimated in the White group, since between families genetical and environmental influences cannot be separated when only data on twins reared together is available. This phenomenon will be interesting in its own right when genetical and environmental effects are estimated.

The magnitude of the between sub-samples component of variation can be estimated from the analysis of variance for each case. The expected mean squares for the between and within samples mean squares for a simple one way analysis of variance are:

Between Samples E.M.S. $= \sigma_w^2 + n\sigma_b^2$

Within Samples E.M.S. $= \sigma_w^2$

where σ_w^2 = variance within samples from the population

σ_b^2 = variance between sub-samples from the population

n = number of observations within each sample

This is true when the number in each sample is the same. Where n differs from sample to sample, then the between pairs expected mean square becomes $\sigma_w^2 + n_0\sigma_b^2$ where $n_0 = (1/(a-1))(N - (\Sigma n_i^2/N))$. In this expression N is the total number of observations, a is the number of samples and n_i are the i different sample sizes. It can be shown (Snedecor and Cochran, 1967) that n_0 is always slightly less than the arithmetic mean of the n_i.

If BMS is the between samples mean square and WMS is the within samples mean square, then unbiased estimates of the two components of variance, $\hat{\sigma}_w^2$ and $\hat{\sigma}_b^2$ are given by:

$$\hat{\sigma}_w^2 = \text{WMS}$$
$$\hat{\sigma}_b^2 = (\text{BMS} - \text{WMS})/N_0$$

The variance component, $\hat{\sigma}_b^2$, was estimated from the formulae given above for each test of each sex-race group. The values of $\hat{\sigma}_b^2$ are given in Table A7. $\hat{\sigma}_b^2$ is an estimate of the magnitude of the between sub-samples component of variance and will be used later in the analysis for interpreting between families genetical and environmental differences. We see that as expected $\hat{\sigma}_b^2$ is larger in Whites than Blacks. It produces a fairly large intraclass correlation in Whites. However, this correlation is negligible in Blacks, as we would expect from the lack of significant differences between the means of the sub-groups in Blacks.

TABLE A7: VALUES OF σ_b^2 AND r FROM THE ANALYSIS OF VARIANCE BETWEEN SUB-GROUPS

	WHITE MALES		WHITE FEMALES		BLACK MALES		BLACK FEMALES	
	σ_b^2	r	σ_b^2	r	σ_b^2	r	σ_b^2	r
Calendar	7.8651	0.1449	0.1877	0.0050	0.3497	0.0112	0.2699	0.0067
Cubes	5.0145	0.0624	1.6606	0.0240	2.5681	0.0458	3.0605	0.0650
Surface Development	43.6915	0.2057	10.7905	0.0957	4.4395	0.0590	-0.9969	0
Vocabulary	1.8187	0.1428	-0.1165	0	0.0416	0.0051	0.3189	0.0354
Form Board	13.0239	0.2091	5.6627	0.1393	0.5862	0.0265	-0.0302	0
Arithmetic	241.9976	0.2863	20.4322	0.0338	-32.9715	0	14.4979	0.0196
Heim	174.1751	0.3389	80.4600	0.2422	71.7612	0.1997	-0.7324	0
Paper Folding	4.2022	0.1622	0.0383	0.0023	-0.1908	0	0.3737	0.0231
Object Aperture	2.8261	0.0614	-0.1935	0	-0.6785	0	1.0736	0.0738
Identical Pictures	286.3841	0.6445	119.9742	0.4370	30.6688	0.1024	27.3525	0.0811
Spelling	49.9608	0.1997	-1.0333	0	-3.2207	0	0.4315	0.0023
Spatial Ability	34.6095	0.0912	26.3707	0.0832	-14.9819	0	-3.1896	0

Note: for negative values of σ_b^2, r has been set to zero since there is no evidence for between groups variability.

SECTION 2: PRELIMINARY ANALYSIS OF THE DATA

1 INTRODUCTION

When any study is undertaken, certain basic assumptions are made from the start in designing the study and collecting the data. If we have tests of all these assumptions, then there is little danger of being seriously mislead. Several assumptions have already been discussed in Section 1. These assumptions are:

1. Random sampling
2. Accurate zygosity diagnosis
3. Homogeneity of the three sub-samples

However, several further assumptions need to be tested as part of the preliminary analysis. These are:

1. That the psychological tests, which were chosen to represent the primary mental abilities identified by Thurstone and others, do in fact represent these abilities. If this assumption fails, other theories of the structure of ability can be considered.
2. That the test scores of the twins are normally distributed. Failure of this assumption may produce biases in many of the analyses reported here.
3. That age does not contribute to variation in the traits being studied. If age is a significant component of variation, the data must be adjusted to remove its effect before further analysis.

The major assumptions underlying the collection of the data and inherent in the preliminary analysis of the data have been mentioned. Different statistical techniques make specific assumptions about the data which will be tested as part of these analyses. When these basic assumptions have been tested, the basic summary statistics will be examined and estimates of the contribution to individual differences of different sources of variation will be made.

2 FACTOR ANALYSIS OF THE TESTS

2.1 Purpose of the Factor Analysis

The tests used in the Georgia Twin Study were chosen to represent the primary mental abilities identified by Thurstone and others and were not just a random selection of tests of ability. They also purport to be free of Jensen's Level I ability and Cattell's g_c (see Section 1.4). However, the actual factor structure of the tests chosen and their inter-relationships will be specific to this study. The factor structure found with this particular combination of tests and the subjects in this study may not conform with findings from previous studies. The Factor analysis will analyse the relationships between the tests in this body of data and arrange them into groups within which the tests are highly correlated. Such groups are known as factors. Factor analysis assumes that the observed measurements (i.e. test scores) and their underlying factors are defined in some fixed population of persons. Therefore, the factors extracted from these data will not necessarily be those extracted from similar bodies of data from other sources. The factors appropriate to this study will be found and the factor structure will be compared with Thurstone's and other models of the structure of ability to see if these tests conform to the original criterion on which they were chosen. The factor structure of the data is important in considering theoretical models of ability and in interpreting the data, since tests belonging to the same factor will be expected to behave in a similar manner.

The usefulness of factor analysis is that it reduces a large number of correlated tests to a small number of uncorrelated factors thus producing a simple structure for interpretation. It also allows some inferences to be made about the common causal relationships between the tests and is, therefore, especially useful for our purpose. The concepts and theory underlying factor analysis are described in Appendix C, which also discusses the two main methods of factor analysing data.

2.2 Methods used in the Factor Analysis

Fourteen variables were included in the factor analysis. These were the thirteen tests described in Section 1.4 and also one derived variable, the number of errors made in the Mazes Test. The analysis was carried out for the total sample of 364 pairs of twins and then separately for Whites (251 pairs) and Blacks (113 pairs). The computations were carried out using a program from the Statistical Package for the Social Sciences (SPSS) (Nie et al, 1970). Principal Factor Analysis with iterations, as described in Appendix C, was used to extract orthogonal factors from the data. Principal Factoring does not produce the maximum-likelihood solution of the factor analytic equation and is dependent on the scaling of $\underset{\sim}{X}$, the matrix of observed measurements. It was used here since the algorithm available for the maximum-likelihood solution of the equation in the SPSS package is poor and does not guarantee convergence. The program uses the matrix of product-moment correlations between the variables and not the variance-covariance matrix. The problem that the fourteen

tests are measured on different scales is thus overcome since the variances of the tests are standardised to unit variance in calculating the correlation coefficients. This is an advantage where unequal total variances are produced by the tests being measured on different scales. However, if we are comparing two or more populations, then reducing the variances to unity may obscure the very population differences we are interested in, if these differences lead to heterogeneity of variances between the populations.

The raw data of the Georgia Twin Study were attached to the Factor Analysis program and the first step the program made was to calculate the 14 x 14 matrix of correlations between the tests. The correlation matrices for the races jointly and then for each race separately are shown in Tables A8a and A8b.

The number of factors to be extracted from the correlation matrix must now be determined. The principal components solution was used to determine the number of factors, where the principal components are defined as exact mathematical transformations of the data and no assumptions about the structure of the tests are made. Usually the first few principal components will account for most of the variance, the remaining variance being test specific. The normal procedure is to extract the same number of factors from the correlation matrix as the number of principal components explaining most of the variation, according to some criterion.

The proportion of variance accounted for by any principal component is λ_i/p where λ_i is the eigenvalue corresponding to the ith component and p is the number of variables. The number of factors extracted from the correlation matrix is usually the same as the number of principal

TABLE A8a: CORRELATION MATRIX OF THE TESTS (UPPER TRIANGLE ONLY) FOR THE TOTAL SAMPLE

	Calendar	Cubes	Arithmetic	Vocabulary	Mazes I	Mazes II	Surface development	Formboard	Heim	Paper folding	Object aperture	Identical pictures	Spatial ability	Spelling
CALENDAR	1000	485	543	454	346	-216	532	506	575	565	414	195	618	524
CUBES		1000	429	352	341	-167	563	488	392	544	462	178	583	350
ARITHMETIC			1000	466	310	-255	532	460	680	502	307	354	587	707
VOCABULARY				1000	245	-191	392	358	639	422	332	229	463	535
MAZES I					1000	109	398	404	333	408	283	340	465	185
MAZES II						1000	-203	-208	-297	-263	-180	-167	-331	-252
SURFACE DEVELOPMENT							1000	618	505	631	541	315	698	424
FORM BOARD								1000	482	615	472	326	667	358
HEIM									1000	531	362	398	616	764
PAPER FOLDING										1000	534	295	730	427
OBJECT APERTURE											1000	184	588	290
IDENTICAL PICTURES												1000	381	228
SPATIAL ABILITY													1000	485
SPELLING														1000

NB: Each entry in the Table must be multiplied by 10^{-3} to get its true value

TABLE A8b:CORRELATION MATRIX OF THE TESTS FOR WHITES AND BLACKS SEPARATELY

(CORRELATIONS FOR WHITES ARE IN THE UPPER TRIANGLE,
CORRELATIONS FOR BLACKS ARE IN THE LOWER TRIANGLE)

	Calendar	Cubes	Arithmetic	Vocabulary	Mazes I	Mazes II	Surface development	Formboard	Heim	Paper folding	Object aperture	Identical pictures	Spatial ability	Spelling
CALENDAR	1000	436	524	377	260	-165	472	400	493	469	364	224	538	447
CUBES	306	1000	404	283	292	-108	540	452	304	496	397	182	525	280
ARITHMETIC	351	225	1000	475	262	-150	503	398	638	465	282	322	524	650
VOCABULARY	356	225	218	1000	185	-087	326	259	628	338	271	274	390	543
MAZES I	214	165	137	077	1000	303	337	322	291	326	209	353	405	124
MAZES II	145	116	267	220	086	1000	-143	-153	-204	-150	-139	-127	-192	-154
SURFACE DEVELOPMENT	365	332	367	225	251	-173	1000	557	436	597	506	326	673	325
FORM BOARD	405	169	283	213	264	-142	431	1000	386	545	415	366	616	233
HEIM	502	275	616	481	114	-315	377	350	1000	436	283	448	519	699
PAPER FOLDING	461	372	298	291	264	-269	436	444	420	1000	502	329	667	326
OBJECT APERTURE	194	373	038	179	155	-121	315	207	222	321	1000	178	548	232
IDENTICAL PICTURES	005	044,	333	027	242	-162	199	124	238	129	070	1000	414	177
SPATIAL ABILITY	494	443	480	300	283	-414	498	423	559	618	416	252	1000	355
SPELLING	449	221	694	343	-005	-248	404	290	793	311	114	210	450	1000

NB: Each entry in the Table must be multiplied by 10^{-3} to get its true value

components having an eigenvalue greater than 1.0. Thus, only components accounting for more than the average variance of a variable will be calculated. The eigenvalues for the fourteen components and their variances as a percentage of the total are presented in Table A9.

It turns out that eigenvalues of greater than 1.0 were obtained for the first three principal components in the case of the total sample and the White group and for the first four principal components in the case of the Black group. This suggests that three factors should be extracted from the correlation matrix in the case of the total sample and the White group and four in the case of the Black group.

The leading diagonal of the correlation matrix was replaced by estimates of communality for each test. The initial estimates of the communalities were given by the squared multiple correlation between each test and the remaining tests. These initial communalities are presented in Table A10. The solution for the number of factors described above was obtained for each group and new communalities were estimated as the multiple correlation with the factors. The procedure was iterated until two consecutive communality estimates did not differ by more than 0.001. Then the factors were rotated using the Varimax Factor Rotation (Kaiser, 1958) available in the SPSS program. This method of rotation was chosen because it produces the closest approximation to simple structure. Varimax rotation attempts to simplify the columns of a factor matrix and defines a simple factor as one with only 0's and 1's in the column. This is equivalent to maximising the squared loadings in each column. The rotated factor matrices for the total sample and for each race are given in Tables A11a to A11c.

TABLE A9: EIGENVALUES AND PERCENTAGE OF VARIANCE ACCOUNTED FOR BY THE FOURTEEN PRINCIPAL COMPONENTS

PRINCIPAL COMPONENT	TOTAL SAMPLE		WHITE GROUP		BLACK GROUP	
	Eigen-value	Percentage of variance	Eigen-value	Percentage of variance	Eigen-value	Percentage of variance
1	6.67111	47.7	5.98666	42.8	5.05500	36.1
2	1.36043	9.7	1.54114	11.0	1.53192	10.9
3	1.06451	7.6	1.27362	9.1	1.22343	8.7
4	0.93728	6.7	0.95765	6.8	1.04857	7.5
5	0.61809	4.4	0.66429	4.7	0.87149	6.2
6	0.54216	3.9	0.54736	3.9	0.78943	5.6
7	0.51216	3.7	0.51789	3.7	0.68804	4.9
8	0.46888	3.3	0.47668	3.4	0.57621	4.1
9	0.42331	3.0	0.42678	3.0	0.51934	3.7
10	0.36036	2.6	0.41722	3.0	0.48088	3.4
11	0.35335	2.5	0.38894	2.8	0.45455	3.2
12	0.28267	2.0	0.31099	2.2	0.31648	2.3
13	0.22568	1.6	0.27378	2.0	0.28453	2.0
14	0.17999	1.3	0.21700	1.6	0.16012	1.1

TABLE A10: INITIAL COMMUNALITY ESTIMATES FOR THE FOURTEEN TESTS

TEST	ESTIMATED COMMUNALITIES BASED ON: TOTAL SAMPLE	WHITE GROUP	BLACK GROUP
Calendar	0.50626	0.42320	0.42380
Cubes	0.43377	0.39986	0.28498
Arithmetic	0.61823	0.58005	0.57059
Vocabulary	0.43435	0.43407	0.28147
Mazes I	0.36399	0.39486	0.25289
Mazes II	0.22760	0.26400	0.25598
Surface Development	0.59112	0.56415	0.38893
Form Board	0.52551	0.45938	0.32237
Heim	0.73568	0.68990	0.73008
Paper Folding	0.60585	0.53356	0.46718
Object Aperture	0.41643	0.37648	0.27955
Identical Pictures	0.27330	0.33687	0.21113
Spatial ability	0.73941	0.67240	0.63996
Spelling	0.67578	0.60895	0.72865

TABLE Alla: VARIMAX ROTATED FACTOR MATRIX FOR THREE FACTORS OF THE TOTAL SAMPLE

	Factor 1	Factor 2	Factor 3
Calendar	477	524	094
Cubes	247	622	092
Arithmetic	699	365	090
Vocabulary	569	303	063
Mazes completed	128	303	1271
Mazes errors	-263	-238	163
Surface development	319	726	114
Form Board	281	681	134
Heim	849	322	103
Paper Folding	334	732	110
Object aperture	168	643	049
Identical pictures	287	251	177
Spatial ability	408	787	134
Spelling	838	204	009

Note: All factor loadings must be multiplied by 10^{-3} to obtain the true values

TABLE A11b: VARIMAX ROTATED FACTOR MATRIX FOR THREE FACTORS OF THE WHITE GROUP

	Factor 1	Factor 2	Factor 3
Calendar	469	460	-017
Cubes	601	219	011
Arithmetic	375	679	-020
Vocabulary	223	637	000
Mazes completed	384	180	1141
Mazes errors	-187	-133	352
Surface development	749	261	002
Form Board	681	191	025
Heim	301	821	030
Paper Folding	716	252	-011
Object aperture	614	126	051
Identical pictures	347	265	148
Spatial ability	806	309	034
Spelling	127	807	-086

Note: All factor loadings must be multiplied by 10^{-3} to obtain the true values

TABLE A11c: VARIMAX ROTATED FACTOR MATRIX FOR FOUR FACTORS OF THE BLACK GROUP

	Factor 1	Factor 2	Factor 3	Factor 4
Calendar	507	300	455	-297
Cubes	166	542	151	-044
Arithmetic	686	124	086	339
Vocabulary	428	255	074	-098
Mazes completed	-056	154	648	203
Mazes errors	-249	-394	196	-187
Surface development	296	465	276	136
Form Board	250	361	428	038
Heim	821	276	075	096
Paper folding	262	580	331	016
Object aperture	-001	574	085	-039
Identical pictures	121	028	139	631
Spatial ability	366	746	212	150
Spelling	882	124	-007	119

Note: All factor loadings must be multiplied by 10^{-3} to obtain the true values

Convergence was extremely slow when the three factor model was used for the total sample and White group and the four factor model for the Black group. For example, convergence required 393 iterations in the case of the total sample. This suggested that something was wrong with our model. From Table A11, we see that the factor loading of Mazes completed on Factor 3 in the Total sample and in Whites is greater than 1.0. Therefore, the test specific variance of Mazes completed is negative. This is a nonsense solution and suggests even more powerfully that the models adopted in the factor analyses are wrong. With this model the first factor accounts for over 65% of the variation in all cases and the first two factors account for more than 86.7% of the variation. This suggests that we may be fitting additional factors which are representing test specific variation rather than variation common to several tests. Looking at Tables A11a to A11c, we see that this is in fact the case. Several tests have high loadings on Factors 1 and 2, but only one test loads highly on additional factors. Only Mazes completed loads on Factor 3 in all cases and the factor loading is greater than 1.0 in the Total sample and the White group. Only identical pictures loads on Factor 4 in the Blacks.

It was decided to repeat the iterative procedure extracting only two factors from the correlation matrix in each case. The final communalities assuming the two factor model are shown in Table A12 and the unrotated factor loadings for this model are shown in Table A13a. Convergence now required only eight iterations for the total sample and the White group and ten iterations for the Black group, which suggests that our two factor model may be more appropriate to these data. The Varimax Rotated Factor Matrices are given in Table A13b.

TABLE A12: FINAL COMMUNALITIES ASSUMING A TWO FACTOR MODEL AND EIGEN-VALUES FOR THE TWO FACTORS

VARIABLE	COMMUNALITIES		
	Total Sample	Whites	Blacks
Calendar	512	432	387
Cubes	452	408	319
Arithmetic	628	603	545
Vocabulary	419	452	228
Mazes completed	268	194	168
Mazes errors	107	043	140
Surface Development	643	626	398
Form Board	571	506	345
Heim	820	750	771
Paper Folding	656	570	535
Object Aperture	420	372	265
Identical Pictures	167	202	053
Spatial Ability	805	756	705
Spelling	763	690	828

GROUP	Factor	Eigen-value	% of variance
Total Sample	1	6.268	86.7
	2	0.963	13.3
White Group	1	5.448	82.5
	2	1.156	17.5
Black Group	1	4.518	79.4
	2	1.169	20.6

Note: The decimal points have been omitted from the Communalities

TABLE A13a:UNROTATED FACTOR LOADINGS FOR THE TWO FACTOR MODEL

	TOTAL SAMPLE		WHITE GROUP		BLACK GROUP	
	FACTOR 1	FACTOR 2	FACTOR 1	FACTOR 2	FACTOR 1	FACTOR 2
CALENDAR	-715	021	-653	071	-617	076
CUBES	-638	-213	-606	-202	-496	270
ARITHMETIC	-740	283	-718	296	-626	-391
VOCABULARY	-604	232	-575	348	-469	-091
MAZES COMPLETED	-472	-214	-404	-175	-273	306
MAZES ERRORS	-309	-108	197	-061	371	051
SURFACE DEVELOPMENT	-767	-233	-746	-263	-607	171
FORM BOARD	-715	-247	-653	-283	-541	228
HEIM	-806	412	-752	430	-790	-383
PAPER FOLDING	-779	-220	-714	-244	-659	316
OBJECT APERTURE	-591	-266	-550	-263	-375	353
IDENTICAL PICTURES	-408	005	-447	-048	-213	-084
SPATIAL ABILITY	-873	-206	-826	-271	-803	244
SPELLING	-695	529	-601	573	-718	-560

NOTE: All decimal points are omitted

TABLE A13b: ROTATED FACTOR MATRICES FOR THE

	TOTAL SAMPLE	
	FACTOR 1	FACTOR 2
CALENDAR	522	489
CUBES	619	263
ARITHMETIC	368	702
VOCABULARY	300	574
MAZES COMPLETED	495	152
MAZES ERRORS	-160	-285
SURFACE DEVELOPMENT	729	332
FORM BOARD	699	288
HEIM	332	842
PAPER FOLDING	730	351
OBJECT APERTURE	619	192
IDENTICAL PICTURES	303	274
SPATIAL ABILITY	791	423
SPELLING	172	857
VARIANCE ACCOUNTED FOR:	3945095	3285550

NOTE: All decimal points are omitted

TWO FACTOR MODEL

WHITE GROUP		BLACK GROUP	
FACTOR 1	FACTOR 2	FACTOR 1	FACTOR 2
456	473	410	467
595	233	191	531
362	687	728	124
219	635	410	244
423	125	000	410
-113	-173	-311	-209
742	276	339	532
683	200	252	530
302	812	845	240
705	270	282	675
591	150	046	513
374	250	215	079
808	320	437	717
095	825	908	059
3688992	2915211	3035134	2650872

The loadings of the tests on the two factors are represented graphically in Figures A1 to A3. The axes are perpendicular since we are attempting to extract orthogonal factors. Oblique rotation allows non-orthogonality of the axes in the factor space in order to obtain the simplest factor structure.

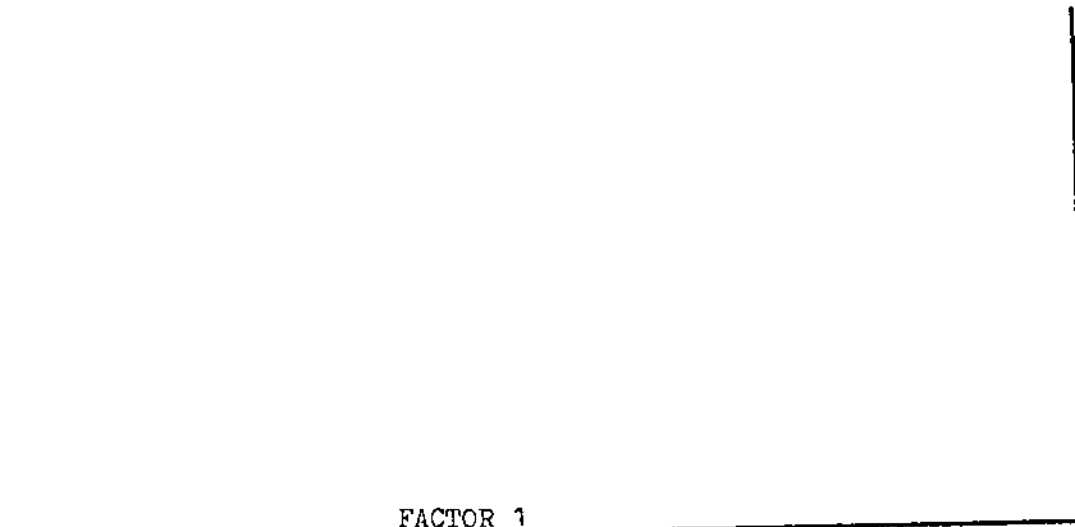
2.3 Discussion and Interpretation of the Factor Analysis

A General Factor

Spearman postulated one of the earliest theories of intelligence, involving two kinds of factor (Spearman 1904). A general factor, g, underlies ability in all intellectual tasks and specific factors, s, produce variation in specific intellectual abilities. In these data, the first principal component can be regarded as an approximation to a general factor accounting for variation common to all tests, analogous to Spearman's g. The first principal component accounts for 47.7 per cent of the variance in the total sample, 42.8 per cent in the White group and 36.1 per cent in the Black group. The remaining components each account for less than eleven per cent of the variance. Thus, a general factor is undoubtedly of some importance in these data, but over 50 per cent of the variation may be accounted for in terms of variation specific to common factors or particular tests.

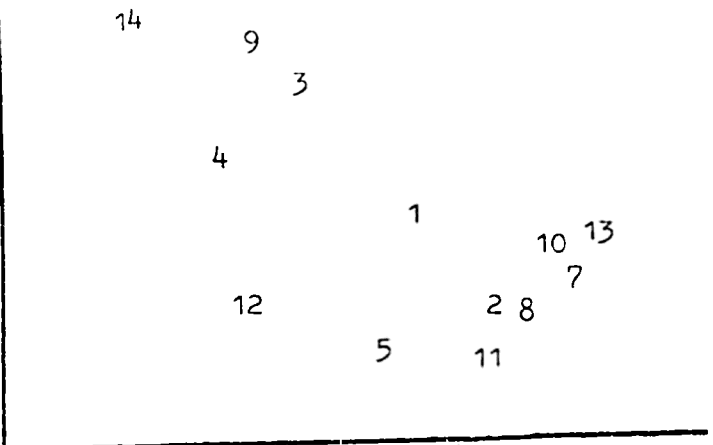
Looking at the communalities of the tests given in Table A12, we can get some idea of the importance of the general factor in performance on these tests. The communalities range from high values for Heim,

FIGURE A1: GRAPHICAL REPRESENTATION OF THE



FACTOR 1

TWO FACTOR MODEL FOR THE TOTAL SAMPLE



KEY:

1. Calendar
2. Cubes
3. Arithmetic
4. Vocabulary
5. Mazes completed
6. Mazes errors
7. Surface development
8. Formboard
9. Heim
10. Paper folding
11. Object aperture
12. Identical pictures
13. Spatial ability
14. Spelling

FACTOR 2

FIGURE A2: GRAPHICAL REPRESENTATION OF THE TWO FACTOR MODEL FOR THE WHITES

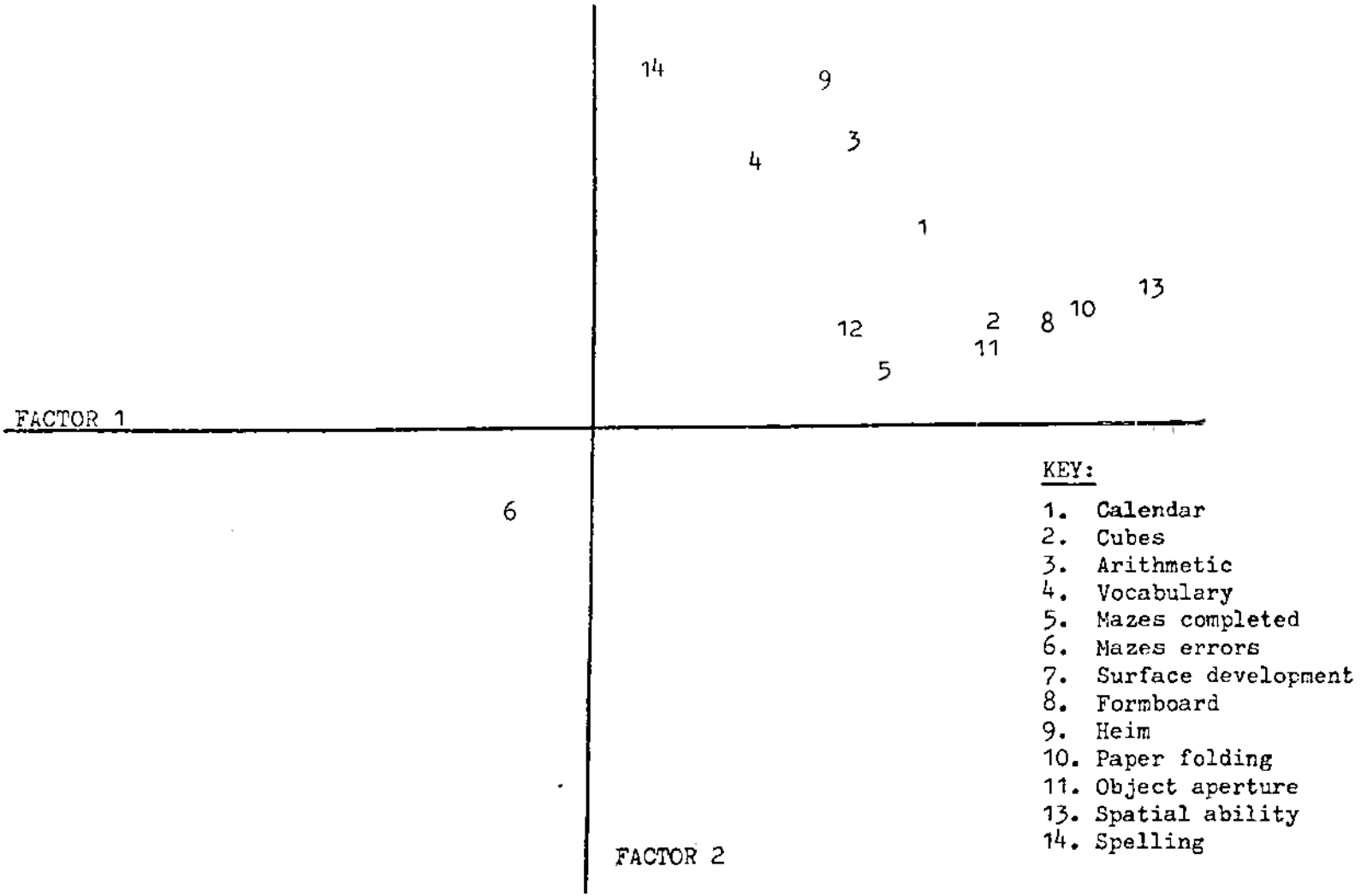
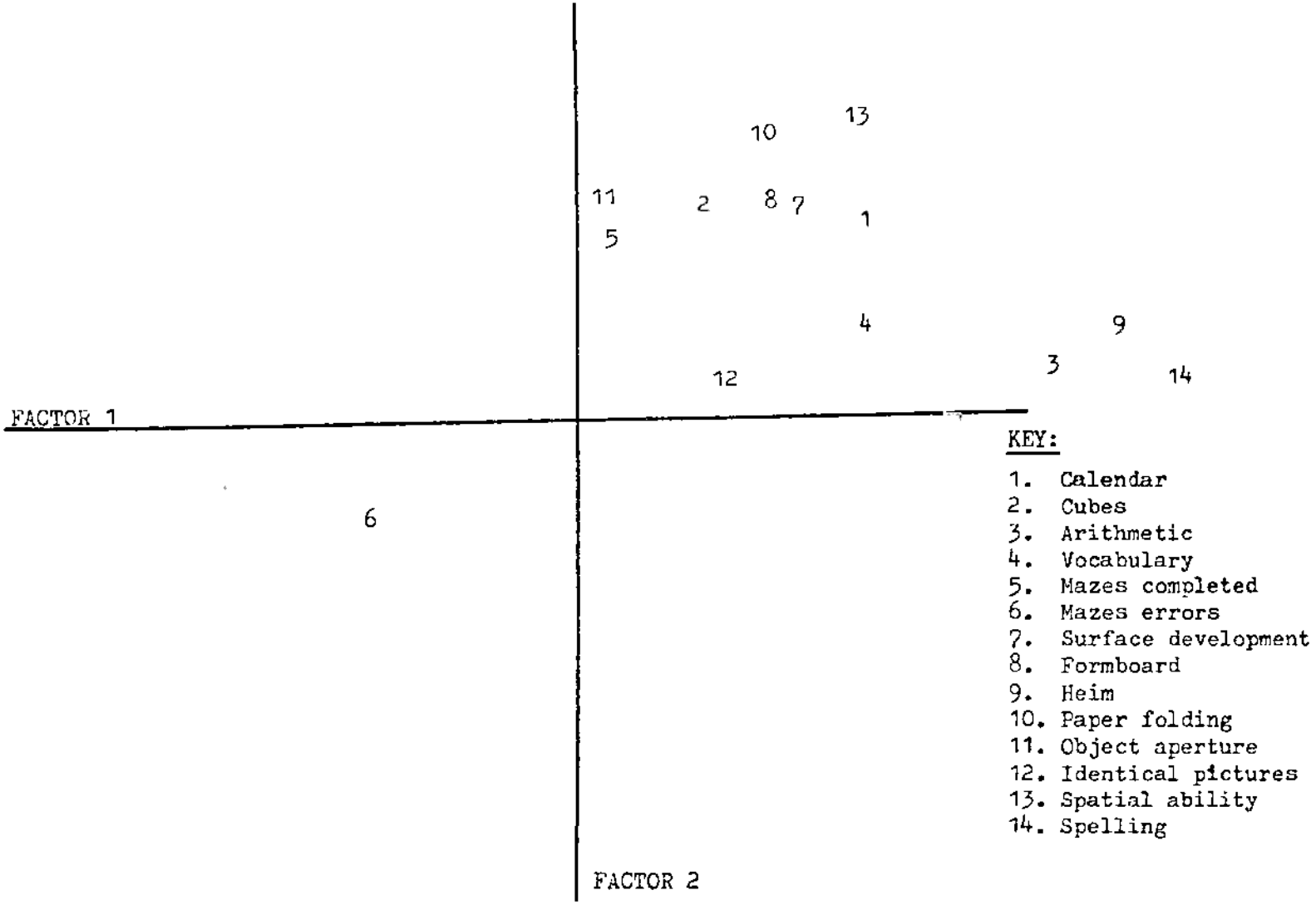


FIGURE A3: GRAPHICAL REPRESENTATION OF THE TWO FACTOR MODEL FOR THE BLACKS



Spatial ability and Spelling to very low values for the Mazes tests and Identical pictures. However, in all cases, the specific variance accounts for a substantial proportion of the variation.

Spearman's two factor theory of ability has been criticised since the general factor does not usually account for all the covariance between the tests. Often the residual covariances between the tests seem to fall into groups on the basis of test content. This suggests another level of factors intermediate between the general factor and the specific factor. Such factors are known as group factors, since they are common to "groups" of similar intellectual activities, but not to all intellectual activities. It is possible that the high level of specific variation found in these data may be attributable to the loading of tests on group factors rather than a general factor and so we shall examine some other theories of performance in the next section.

The extremely low communalities of the Mazes tests and Identical pictures test are worrying. It seems that virtually all variation in performance on these tests is test specific, which may be an indication of the low reliability of the tests. If these tests do have a low reliability, the within pairs variance for these tests will be inflated. We will look at the possibility later.

The Factor Structure of these Tests

Thurstone (1938) identified a number of broadly based group factors. According to his theory Spearman's general factor is of only minor importance in determining test performance. However, performance in

particular tasks could be predicted if performance on his group factors, which he called primary mental abilities, were known.

The battery of tests used in this study were chosen to represent four of Thurstone's primary mental abilities:

1. Verbal comprehension - V
2. Number - N
3. Space - S
4. Perceptual speed - P

We shall now see how far the tests chosen do in fact represent these four primary mental abilities.

It was impossible to extract four meaningful factors from the data. Whenever more than two factors were extracted, only one test loaded on the additional factor. This was always one of the Mazes tests or the Identical pictures test. This suggested that additional factors after the first two are redundant since they represent only test specific variation and not variation common to a group of tests. Thus, attempts at extracting Thurstone's four primary mental abilities were abandoned and the two factor model was adopted. The final categorisation of the tests into two factors is shown in Table A14. Each test is listed under the factor on which it had the highest loading. The two factors were easily identifiable. One clearly measured Spatial ability and the other Verbal-Arithmetic ability.

The first factor accounts for the largest proportion of the variance. In the Total sample and the White group, the first factor

TABLE A14: SUMMARY OF THE TWO FACTORS

Factor 1 Spatial	Factor 2 Verbal-Arithmetic
Spatial ability	Spelling
Paper Folding	Heim
Surface development	Arithmetic
Form Board	Vocabulary
Cubes	
Object aperture	
Calendar	
Mazes completed	
Identical pictures	
Mazes errors	

was the Spatial factor. However, in the Black group the first factor was the Verbal-Arithmetic factor. Thus although the pattern of loadings on the factors, which were identifiable as Spatial and Verbal-Arithmetic factors was similar in the two races, their relative importance in causing variation was not. We must bear this in mind in later sections when we are looking for the source of individual differences, in terms of the variance components.

From Table A13b, we see that although the Varimax criterion was used to obtain a solution, the approximation to simple structure is poor when the factors are constrained. i.e. many of the loadings in each column are intermediate. This provides some problems in interpretation. The Calendar test loaded highest on the Spatial factor in all cases and is listed under this factor in Table A14. However, its loading on this factor is intermediate and not very different to its loading on the Verbal-Arithmetic factor. From Appendix B, we can see that the material in this test is not similar to that of other tests of the two factors and we expect that this test will not prove easy to interpret.

Most other tests have a reasonably high loading on one factor and low loading on the other factor and may be grouped easily. However, the Mazes tests and the Identical pictures have low loadings on both factors. They are all listed under the Spatial factor since the test material is clearly spatial. However, these tests will probably be difficult to interpret. We have already noted that these tests may have a low reliability and their low loading on both factors is a further indication of this.

Thus the Factor Analysis does not show that the tests selected for this study represent four of Thurstone's primary mental abilities. Broadly they may be considered to represent his Space and Verbal Comprehension factors, but the inclusion of Arithmetic with the verbal tests is clearly an anomaly.

More recent theories of ability have suggested a two factor model for intelligence. Cattell's crystallised and fluid intelligence and Jensen's Level I and Level II abilities were described in Section 1.4. Since we have identified two factors, these models may be more appropriate. Osborne (personal communication) points out that these tests were originally selected to be free of Level I type abilities. However, Level I abilities and g_c both depend on past experience and educational achievements and are, thus, broadly similar and it seems unlikely that tests of Verbal ability or Arithmetic could ever be entirely independent of past education. It seems likely that tests of the Verbal-Arithmetic factor might load quite heavily on Jensen's Level I ability or Cattell's g_c. It has been claimed, however, that Spatial tests are culture-free and do not depend on past experience. Thus, such tests may more truly represent Level II type abilities and g_f, which involve mental manipulation of new ideas.

We see that these data are consistent with the interpretation that the two factors extracted represent Jensen's Level I and Level II abilities (which are similar to Cattells g_c and g_f), since we have extracted two factors one of which appears to represent tasks which rely on education and past experience, and the other tasks where past experience is not essential. However, although these data appear to be inconsistent with Thurstone's theory of primary mental abilities and

more consistent with Jensen's theory of Level I and Level II abilities, this could be an artefact of the tests chosen. We notice that Spatial tasks are relatively over-represented, at the expense of Verbal and Arithmetic tasks, especially Arithmetic tasks. Therefore, the loading of the Arithmetic on a primary Verbal factor is not such an anomaly as it first seems, since this test has to load on one or other factor. We have ten Spatial tests, which all load on one Spatial factor. Thurstone's theory predicts that these tests should load on two factors - Space and Perceptual Speed - but once again we see an imbalance in the tests chosen since only two tests (i.e. Mazes errors and Identical pictures) out of the eight were expected to load on Perceptual Speed. Therefore, we accept that the two factors theory of ability is the most appropriate for this body of data, but that this does not necessarily imply that Thurstone's model would not be appropriate given a more representative battery of tests.

In this thesis, we will report analysis of:

1. each individual test of the battery
2. a "general" factor extracted from the data.

However, since the tests may be considered representative of two factors, it would be a useful addition to the work already done to obtain and analyse factor scores for the two factors. This analysis will be performed and reported at a later date.

3 BASIC STATISTICS DERIVED FROM THE RAW DATA

Basic statistics were calculated from the raw data, separately by race, sex and zygosity, using a programme from the Statistical Package for the Social Sciences (SPSS). The statistics calculated i.e. the means, standard errors, variances, standard deviations, ranges and measures of skewness and kurtosis, will be discussed in the following sections.

3.1 Skewness and Kurtosis

Measures of many psychological and biological attributes are assumed to be distributed normally. Such measures vary continuously. That is, theoretically at least, the score of an individual can take any value within the range of variation. However, the scores of most individuals lie within the central portion of the distribution. It will be assumed that the test scores of the individuals of the Georgia Twin Study are normally distributed through much of the analysis and this assumption will be tested here. Data can deviate from normality in two main ways:

Skewness

If there are too many individuals with scores above the mean or conversely if there are too many individuals at the lower end of the distribution, then the distribution is said to be skewed (positively skewed in the first case and negatively skewed in the latter case). A

measure of the amount of skewness is given by the coefficient of skewness which is the average value of $(X-\mu)^3/\sigma^3$ taken over the population, where X is the test score, μ is the population mean and σ^2 is the population variance. This is the third moment about the mean. Division by σ^3 ensures that the coefficient is independent of the scale of measurement. If the data are normally distributed, the coefficient of skewness is also normally distributed with a mean of zero and a standard deviation of $\sqrt{6/N}$. Snedecor and Cochran (1967) give tables of the one-tailed 5% and 1% significance levels of the coefficient of skewness. The coefficients of skewness for the Georgia Twin Study are given in Table A15, together with their significance levels.

Kurtosis

Data may deviate from normality by showing kurtosis. A measure of kurtosis is given by $(X-\mu)^4$ divided by σ^4 over the population i.e. the fourth moment about the mean. When kurtosis is positive, too many individuals have scores either near the mean or far away from it and the curve appears more pointed than usual. In negative kurtosis, the curve appears flatter than usual. The expected value of the coefficient of kurtosis is 3.0 and Snedecor and Cochran (1967) tabulate, for the two tails, 5% and 1% significance levels. The coefficients of kurtosis for the Georgia Twin Study are given in Table A16, together with the corresponding significance levels.

It can be seen that there are deviations from normality in these data which, however, are by no means consistent over all tests or over

TABLE A15: COEFFICIENTS OF SKEWNESS FOR THE EIGHT GROUPS OF TWINS ON THE FOURTEEN TESTS

	WHITES				BLACKS			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
NUMBER IN GROUP	132	84	150	136	44	24	94	64
ARITHMETIC	178	-129	148	270	689†	516	481†	079
VOCABULARY	600*	385	441†	518*	-014	568	931*	534†
HEIM	-311	-076	-158	-118	138	563	504†	463
SPELLING	-485†	-171	-850*	-807*	001	358	-068	-201
CALENDAR	-045	358	-191	299	053	522	1256*	313
IDENTICAL PICTURES	845*	488†	640*	850*	532	684	751*	359
CUBES	411†	539†	407†	631*	355	720†	640*	289
SURFACE DEVELOPMENT	259	708*	780*	661*	1182*	1596*	636*	567†
FORMBOARD	718*	381	648*	641*	771†	650	1092*	940*
PAPER FOLDING	-140	-002	014	-125	525	524	834*	662†
OBJECT APERTURE	335	887*	449†	658*	600†	884†	391	003
SPATIAL ABILITY	-251	-349	-076	-069	256	865†	431†	259
MAZES COMPLETED	170	309	531*	-033	503	527	1104*	1255*
MAZES ERRORS	1688*	1729*	1961*	1129*	2120*	553	1280*	1226*

NOTE: All figures in the Table must be multiplied by 10^{-3} to obtain the true coefficients

KEY: † Significant at the 10% level
* Significant at the 2% level

TABLE A16: COEFFICIENTS OF KURTOSIS FOR THE EIGHT GROUPS OF

	WHITES			
	MALES		FEMALES	
	MZ	DZ	MZ	DZ
NUMBER IN GROUP	132	84	150	136
ARITHMETIC	3.655	2.822	2.510	2.885
VOCABULARY	3.351	3.650	2.901	3.107
HEIM	2.355*	2.353	2.354*	2.349*
SPELLING	2.281*	2.269*	2.893	2.899
CALENDAR	3.897*	2.559	2.973	2.733
IDENTICAL PICTURES	3.232	2.643	2.941	3.908*
CUBES	3.084	2.818	3.267	3.626
SURFACE DEVELOPMENT	1.988**	3.139	3.402	3.036
FORMBOARD	2.788	2.320	2.871	3.154
PAPER FOLDING	2.919	3.230	3.134	2.406*
OBJECT APERTURE	2.234**	3.639	2.947	3.772*
SPATIAL ABILITY	2.493	2.467	2.665	2.335*
MAZES COMPLETED	2.718	2.216*	2.592	3.243
MAZES ERRORS	6.835**	6.891**	8.419**	4.050**

KEY: * Significant at the 5% level
** Significant at the 1% level

TWINS

BLACKS

MALES		FEMALES	
MZ	DZ	MZ	DZ
44	24	94	64
2.756	2.176	3.014	1.811**
3.205	2.447	4.261*	3.040
2.313	2.127	2.638	2.524
1.722**	2.210	1.882**	1.880**
3.420	2.806	5.467**	2.659
4.185*	2.616	2.850	3.783
2.372	2.862	3.485	2.800
4.435*	6.955**	3.536	3.958*
2.815	2.453	4.557**	3.493
2.318	2.525	2.790	3.302
3.417	3.607	2.858	2.704
2.012*	3.459	2.084**	3.470
3.216	2.307	3.749	4.187*
6.769**	2.405	3.259	3.394

all groups. Such deviations from normality may be due to a number of causes. The most obvious is that non-random sampling may produce a sample which is non-normal although it has been drawn from a normally distributed population. However, this is not the only cause and the true variability in a population may not be normally distributed due to genetical and/or environmental effects which are influencing the traits being studied. For example, if a trait is determined by only a small number of genes, then dominance may produce a skewed distribution. Similarly epistasis, genotype-environment interaction and other non-additive properties of the genes may produce skewness. The problem of non-additivity and genotype-environment interaction will be further discussed in a later Section.

Most of the techniques used in analysis of this data are robust to deviations from normality of the magnitude found here. However, the Mazes errors Test shows gross deviations from normality, and the coefficients of skewness and kurtosis deviate by a large amount in the positive direction from the expectations on the basis of normality, suggesting that this test could be measured on a better scale and that an interpretation of the genetical and environmental influences affecting the trait will be difficult on the present scale. A simple logarithmic transformation normalised the distribution of the Mazes errors scores and it was decided to replace the Mazes errors scores by log mazes errors scores in all analyses.

3.2 The Means and Variances

The mean ages and test scores, calculated separately by race, sex and zygosity, are presented in Table A17a. It can be seen that the mean ages and test scores of the eight groups differ. The significance and interpretation of these mean differences are discussed in the next section. All other summary statistics calculated by the SPSS program are derived from the total variances of the eight groups. Therefore, these variances calculated separately by race, sex and zygosity are given in Table A17b. The total variances of the eight groups differ and these differences seem to be systematic. Whites are more variable than Blacks and males more variable than females. The significance of these differences and their underlying causes will be investigated in later sections.

3.3 Simple Analysis of Variance of the Data

The total variances were divided into the between pairs mean square and the within pairs mean square by a simple one-way analysis of variance (Snedecor and Cochran, 1967). These mean squares are presented in Table A18a. The significance levels of the between pairs mean square when tested against the within pairs mean square is given in Table A18b. We see that in nearly all cases there were significant differences between pairs. Those cases where the between pairs mean square was not significant are generally those where the sample size was small and, therefore, the power of the test for detecting a significant between pairs mean-square was reduced. The sources of

TABLE A17a: MEANS OF THE GROUPS (not corrected for differences in mean age between the groups)

	WHITES				BLACKS			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
AGE	15.712	15.429	15.480	15.647	14.636	15.000	14.915	15.187
ARITHMETIC	70.341	63.667	67.120	66.699	41.545	31.958	48.894	49.687
VOCABULARY	5.826	4.440	5.087	4.757	3.523	2.125	2.298	2.406
HEIM	43.288	38.583	42.153	36.824	27.295	21.187	24.255	25.547
SPELLING	34.235	30.631	37.773	37.140	22.205	16.250	26.649	25.703
CALENDAR	14.644	11.774	13.187	13.007	8.023	5.542	7.064	6.672
IDENTICAL PICTURES	57.295	55.369	56.113	55.059	48.205	46.708	55.543	50.312
CUBES	10.023	8.202	7.107	7.228	2.364	4.250	0.894	0.078
SURFACE DEVELOPMENT	27.932	23.607	25.080	23.309	15.659	14.833	16.319	15.031
FORMBOARD	14.530	11.619	11.173	11.221	5.977	7.208	5.532	5.312
PAPER FOLDING	8.621	7.774	7.820	7.493	3.545	4.000	3.234	2.297
OBJECT APERTURE	7.258	6.143	3.607	3.074	2.205	1.292	0.266	- 0.781
SPATIAL ABILITY	61.682	58.940	53.040	51.669	40.523	38.000	33.255	30.891
MAZES COMPLETED	31.114	30.464	28.627	27.059	23.866	22.677	22.489	21.687
MAZES ERRORS	0.832	0.896	0.856	0.904	0.864	0.812	1.080	1.002

TABLE A17b: TOTAL VARIANCES OF THE EIGHT GROUPS OF TWINS (not corrected for the regression of test score on age)

	WHITES				BLACKS			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
NUMBER OF PAIRS	66	42	75	68	22	12	47	32
ARITHMETIC	630.913	723.839	691.341	781.427	855.882	455.868	711.967	850.218
VOCABULARY	12.603	11.599	14.523	12.615	6.441	7.592	9.416	9.166
HEIM	369.214	376.897	343.849	342.354	342.073	347.101	315.848	258.918
SPELLING	205.998	189.127	167.472	183.662	228.957	177.500	258.080	211.609
CALENDAR	54.170	37.647	37.146	43.118	37.232	24.955	47.759	30.065
IDENTICAL PICTURES	238.897	225.368	243.296	197.108	225.143	239.955	311.907	305.996
CUBES	81.305	87.368	68.713	73.614	66.888	47.761	59.838	24.803
SURFACE DEVELOPMENT	197.621	179.133	116.947	111.608	77.486	73.101	43.854	54.539
FORMBOARD	57.259	37.540	40.856	41.269	19.790	18.085	15.542	13.583
PAPER FOLDING	21.993	20.852	18.524	14.770	18.812	11.826	18.568	13.006
OBJECT APERTURE	47.994	50.991	29.690	23.846	23.887	37.694	15.703	11.062
SPATIAL ABILITY	340.066	369.960	302.401	351.145	386.674	194.435	245.439	114.893
MAZES COMPLETED	68.865	93.818	97.578	66.841	64.382	89.884	85.951	85.806
MAZES ERRORS	0.190	0.189	0.167	0.151	0.211	0.135	0.182	0.238

TABLE A18a: BETWEEN PAIRS AND WITHIN PAIRS MEAN SQUARES

	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
Arithmetic	1114.608 128.917	997.808 382.595	1059.392 123.853	966.582 251.434	1621.510 133.773	786.646 217.375	1334.444 103.128	1489.603 264.750
Vocabulary	18.342 4.977	15.298 6.560	19.957 5.620	11.406 8.610	9.198 4.205	6.690 4.345	12.424 5.447	11.549 6.906
Heim	659.653 52.318	525.603 203.798	488.016 49.713	426.291 94.646	599.055 79.159	520.511 169.167	558.242 69.319	381.950 108.266
Spelling	382.936 21.386	261.348 101.036	236.390 22.693	237.183 57.640	469.730 20.447	282.247 101.500	441.787 59.564	345.197 77.547
Calendar	83.188 25.235	52.003 24.250	48.025 19.920	51.804 21.404	56.346 18.068	37.056 14.542	72.060 20.702	44.054 17.484
Identical pictures	405.598 65.705	354.695 104.274	417.116 55.180	285.989 85.618	306.473 144.386	248.323 229.125	482.518 152.096	478.887 147.156
Cubes	130.116 32.850	132.567 43.060	87.134 45.653	87.358 53.390	78.498 43.000	50.540 46.417	81.778 40.043	27.756 22.453
Surface development	344.561 33.295	275.420 76.393	185.501 33.573	136.632 78.824	115.320 35.614	125.831 32.083	63.280 21.043	65.290 41.531
Formboard	93.116 21.439	59.501 17.429	56.744 21.627	53.326 27.191	31.068 9.614	29.686 9.792	13.670 14.340	16.179 10.906
Paper folding	34.422 9.424	30.500 12.155	25.108 7.713	20.566 7.728	26.812 6.727	10.938 9.333	28.497 7.799	14.142 12.016
Object aperture	91.550 22.030	76.394 27.929	40.014 14.340	31.411 16.029	26.823 9.523	59.506 19.958	20.632 9.500	13.615 8.544
Spatial ability	593.359 84.742	615.033 187.655	469.756 55.133	513.359 152.375	670.370 57.295	518.478 96.167	453.757 32.447	174.199 62.703
Mazes completed	96.838 41.114	153.725 38 917	158.440 33.547	86.981 44.365	89.423 20.068	127.071 62.167	121.542 53.213	94.299 69.406
Mazes errors	0.817 0.069	0.160 0.117	0.192 0.047	0.152 0.076	0.241 0.072	0.054 0.114	0.212 0.097	0.238 0.102

TABLE A18b: SIGNIFICANCE OF THE BETWEEN PAIRS MEAN SQUARE

	WHITES				BLACKS			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
ARITHMETIC	***	**	***	***	***	*	***	***
VOCABULARY	***	**	***	0.13	*	0.24	**	†
HEIM	***	**	***	***	***	*	***	***
SPELLING	***	**	***	***	***	*	***	***
CALENDAR	***	**	***	***	**	†	***	**
IDENTICAL PICTURES	***	***	***	***	*	0.44	***	***
CUBES	***	***	**	*	†	0.44	**	0.28
SURFACE DEVELOPMENT	***	***	***	*	**	*	***	0.11
FORMBOARD	***	***	***	**	**	*	0.56	0.14
PAPER FOLDING	***	**	***	***	**	0.39	***	0.32
OBJECT APERTURE	***	***	***	**	*	*	**	†
SPATIAL ABILITY	***	***	***	***	***	*	***	**
MAZES COMPLETED	***	***	***	**	***	0.12	**	0.20
MAZES ERRORS	***	0.16	***	**	**	0.88	**	*
NUMBER OF PAIRS	66	42	75	68	22	12	47	32

KEY: *** significant at the 0.1% level
** significant at the 1% level
* significant at the 5% level
† significant at the 10% level

NOTE: The figures given in the body of the Table are probabilities given in decimal form for those between pairs mean squares which are not significant

variation leading to significant differences between pairs will be examined in later Sections. However, before we can continue the analysis, there is one further assumption that must be tested.

4 AGE CORRECTION

The ages of the subjects in this study varied from twelve to nineteen years. Covariance between age and test score is a source of variation which will produce biases if it is not recognised in the analysis. Covariance between age and test score will produce a regression of test score on age within groups. Mean differences in age between the groups will then produce mean differences in test score between groups purely as a function of age. We have seen from Table A17a that there are mean differences in both age and test score between the groups, although we have not tested their significance; and we would like to know the source of these differences.

Covariation between age and test score will lead to the regression of test score on age and hence to an increase in the total variances. The most obvious expectation is that age is a simple between pairs effect, leading to an increase in the between pairs mean squares and the regression of between pairs variance on age. We have no a priori expectation for a regression of within pairs variance on age; however, such a regression, if found, could be interpreted as a within families effect or some type of interaction between age and the determinants of intra-pair differences.

4.1 Analysis of Covariance

In an analysis of covariance differences in mean age and test score between groups are tested for significance. The significance of any regression of test score on age is then found. The mean test scores

are adjusted to remove the age effect and tested to see if significant mean differences in test score remain.

The analysis of covariance combines features of regression analysis and the analysis of variance and is discussed by Snedecor and Cochran (1967). The computations described below were carried out for each of the tests separately.

Firstly between and within groups sums of squares for age, X_{ij}, and test scores, Y_{ij}, were calculated in the usual way,

where X_{ij} is the deviation of ith individual in the jth group from the mean age of the sample

and Y_{ij} is the deviation of the ith individual in the jth group from the mean test score of the sample.

The next step was to compute the total and between groups sums of products as follows:

$$\text{Total Sum of Products} = \sum_{ij} X_{ij}Y_{ij}/j - \left(\sum_{ij} X_{ij} \sum_{ij} Y_{ij}\right)/j$$

$$\text{Between groups Sum of Products} = \sum_{j}\left(\sum_{i} X_{ij} \sum_{i} Y_{ij}\right) - \left(\sum_{ij} X_{ij} \sum_{ij} Y_{ij}\right)/j$$

where i is the number in a group

j is the number of groups

The within groups sums of products were found by difference. The Analysis of Covariance Table, given for the case of the Arithmetic Test in Table A19, was then constructed. The between groups mean squares for age and test score were tested against their appropriate within groups mean squares. The between pairs mean squares for age and test scores were highly significant for all fourteen tests. The next step was to determine how far the significant mean differences in test score could be explained by mean differences in age. The regression of test score on

TABLE A19: ANALYSIS OF THE COVARIANCE BETWEEN SCORES ON THE ARITHMETIC TEST AND AGE

A. The Basic Analysis of Covariance Table

Source	df	SS (x)	SP (x,y)	SS (y)
Total	727	1727 . 8407	10432 . 1951	599210 . 5371
Between groups	7	76 . 4588	2289 . 7249	80915 . 1780
Within groups	720	1651 . 3818	8142 . 4702	518295 . 3590

B. Test of the Regression of Test Score on Age

Source	df	Sum of Squares	Mean Squares	F Ratio
Reduction due to regression	1	40148 . 0874	40148 . 0874	60 . 3715
Deviations from regression	719	478147 . 2725	665 . 0170	
Within groups	720	518295 . 2725	719 . 8545	

C. F Ratios for Testing Significant Differences Between the Means

	F Ratio	df	Significant Level
X means	4 . 7623	(7,720)	* * *
Y means	16 . 0578	(7,720)	* * *
Y* means	12 . 4759	(7,719)	* * *
B	60 . 3715	(1,719)	* * *

Regression Coefficient = 4 . 9307

KEY

X = Age
Y = Test score
Y* = Test score after age adjustment.

age within groups was tested for significance. This test is done by firstly dividing the within groups sums of squares into two parts: one part due to the regression and one part due to deviations from the regression. The regression sum of squares was calculated as follows:

Regression sum of squares = $W_{XY}2/W_{XX}$

where W_{XY} = the within groups sum of products

W_{XX} = the within groups sum of squares

It has one degree of freedom. The deviation sum of squares was found by difference. The regression mean square was tested against the deviation mean square. This is shown in the Part B of Table A19.

The regression was significant for all tests. The regression coefficient was calculated as W_{XY}/W_{XX} and the means were adjusted as follows:

$$\overline{Y}_j^* = \overline{Y}_j - b\,(\overline{X}_j - \overline{X}_{ij})$$

where $\overline{X}_{ij}$ is the overall mean. The adjusted means for the eight groups of twins for all fourteen tests are given in Table A20. The only remaining task was to test for significant differences among the adjusted means. The null hypothesis that the adjusted means are equal was tested. The method is given by Snedecor and Cochran (1967; pp 424-425). The age-adjusted means differed significantly for every test. The course of the mean differences will be investigated in Section 2.5.

4.2 Multiple Regression Analysis

We have already mentioned that covariance between age and test score leads to an increase in the total variance, which can probably be

TABLE A20: MEANS OF TWIN GROUPS (corrected for mean differences in age between the groups)

TEST	CAUCASIANS				NEGROES			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
ARITHMETIC	67.9715	62.6531	65.8513	65.0348	44.9272	32.9974	50.3404	44.6600
VOCABULARY	5.4746	4.2895	4.9003	4.5102	3.9356	2.2792	2.5125	3.2330
HEIM	41.2514	37.7110	41.0711	35.3936	29.6863	22.0804	25.4982	23.7656
SPELLING	32.8797	30.0567	37.0530	36.1881	23.7963	16.8445	27.4760	22.3013
CALENDAR	14.1916	11.5803	15.2396	12.6140	8.5542	5.7405	7.3401	7.4544
IDENTICAL PICTURES	56.5161	55.0355	55.6992	54.5119	49.1196	47.0497	56.0198	51.0024
CUBES	9.6207	8.0297	6.8933	6.9454	4.8364	4.4265	1.1396	1.3641
SURFACE DEVELOPMENT	27.1367	23.2665	24.6575	22.7504	16.5928	15.1819	16.8045	14.2988
FORMBOARD	14.2347	11.4926	11.0161	11.0136	6.3237	7.3375	5.7123	4.6966
PAPER FOLDING	8.3291	7.6490	7.6649	7.2880	3.8878	4.1281	3.4122	3.8332
OBJECT APERTURE	6.9001	5.9898	3.4169	2.8227	2.6252	1.4490	0.4844	0.6595
SPATIAL ABILITY	60.4575	58.4155	52.3893	50.8086	41.9613	38.5374	34.0028	35.0904
MAZES COMPLETED	30.7344	30.3015	28.4253	26.7924	24.3118	22.8435	22.7207	21.8413
MAZES ERRORS	8.8931	10.1692	9.0105	9.6792	9.5712	6.6894	16.3157	11.2550

ascribed to an increase in the between families variance rather than the within families variance. Therefore, before further analysing the variances, we will correct them for any regression of test score on age, using the generalised multiple regression analysis described in Appendix D. In our case, we have only one predictor variable, namely age, but a series of fourteen criterion variables. Thus, following the notation of Appendix D, the full model may be written as:

$$
\begin{aligned}
x_{ij} &= \beta_{1,0} z_{0N} + \beta_{1,1} z_{1N} + e_{1,N} \\
x_{2j} &= \beta_{2,0} z_{0N} + \beta_{2,1} z_{1N} + e_{2,N} \\
&\vdots \\
x_{14,j} &= \beta_{14,0} z_{0N} + \beta_{14,1} z_{1N} + e_{14,N}
\end{aligned}
$$

Now we may define the following matrices:

$$
x_1 = \begin{bmatrix} x_{1,1} \\ \vdots \\ x_{1,N} \end{bmatrix} \qquad x_2 = \begin{bmatrix} x_{2,1} \\ \vdots \\ x_{2,N} \end{bmatrix} \qquad x_{14} = \begin{bmatrix} x_{14,1} \\ \vdots \\ x_{14,N} \end{bmatrix}
$$

and let:

$$
X = [x_1 \; x_2 \; \ldots\ldots\ldots\ldots \; x_{14}]
$$

An analogous notation is used for $e_1 \; e_2 \; \ldots\ldots \; e_{14}$ and E.

Now

$$
Z = \begin{bmatrix} z_{01} & z_{02} & \ldots\ldots\ldots & z_{0,N} \\ z_{11} & z_{12} & \ldots\ldots\ldots & z_{1,N} \end{bmatrix}
$$

Let us now define the matrices of coefficients:

$$\beta_1 = \begin{bmatrix} \beta_{10} \\ \beta_{11} \end{bmatrix} \quad \beta_2 = \begin{bmatrix} \beta_{20} \\ \beta_{21} \end{bmatrix} \cdots\cdots\cdots \beta_{14} = \begin{bmatrix} \beta_{14,0} \\ \beta_{14,1} \end{bmatrix}$$

so that:

$$B = [\beta_1 \quad \beta_2 \quad \cdots\cdots\cdots \beta_{14}]$$

The full model, Ω , is now written as:

$$\underset{(Nx14)}{X} = \underset{(Nx2)}{Z'} \quad \underset{(2x14)}{B} + \underset{(Nx14)}{E}$$

This equation may now be solved for B as described in Appendix D. This full model is designated Ω by Seal (1964). Now a reduced model, ω , is formulated. In this model B_1 is eliminated from B and the corresponding row of Z is also eliminated. The residuals from the Ω are tested against the residuals from the ω model to see if fitting the Ω model produces a significant reduction in the residuals (see Appendix D). If the chisquare is significant, then the effect of β_1, i.e. age is considered to be significant.

The calculations for this analysis were performed using a program written by Dr. L.J. Eaves of the Genetics Department of the University of Birmingham.

The program firstly constructs the matrix of sums of squares and sums of products among all the variables, separately by sex, race and zygosity, and then performs the regression analysis and tests of significance as described in Appendix D.

The analysis was performed using the between pairs sums of squares and sums of products matrix, calculated from the pair sums.

This analysis tests for the regression of between pairs of differences on age. Then the analysis was repeated using the within pairs sums of squares andsums of products matrix, based on the pair differences, which tests for the regression of within pairs differences on age. The results of the analyses are presented in Tables A21 and A22. The multiple regression analysis was performed on the matrices of sums of squares and sums of products derived from the raw scores for all tests, including Mazes errors. Strictly the analysis should have used the log Mazes errors scores. However, although they were not used we do not expect that the regression coefficients will be seriously biased. The log Mazes errors Mean Squares were age-corrected in a separate analysis.

We can see that there is little evidence for the regression of within pairs differences on age, although the regression is significant at the five per cent level for monozygotic Black males. However, the regression of between pairs variance on age was significant for the two largest groups, the monozygotic and dizygotic White females and approached significance in the case of the monozygotic Black females. Thus it seems that the regression of test score on age found during the Analysis of Covariance is due to the regression of the between pairs variance on age. If no adjustment is made for this regression, then the between pairs variance will be inflated compared with the within pairs variance and between families genetical and environmental components of variation will be over-estimated. Therefore, the total variances and between families variances must be corrected for the effects of age. During the analysis, the matrix of regression sums of squares and sum of products was calculated. The diagonal terms of

TABLE A21: SUMS OF SQUARES DUE TO THE REGRESSION OF BETWEEN PAIRS DIFFERENCES IN TEST SCORE ON AGE

	WHITES				BLACKS			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
ARITHMETIC	2806.26	4097.34	16385.30	24600.80	1029.67	10.00	1325.89	403.61
VOCABULARY	148.58	75.28	285.48	364.67	0.52	55.23	93.90	9.96
HEIM	2692.25	1698.86	11879.80	11644.60	986.58	748.23	1041.93	1399.68
SIEMLING	1066.34	1000.17	5994.86	5220.71	6.61	42.03	1321.48	493.94
CALENDAR	106.74	26.09	534.96	946.40	76.55	28.90	225.93	12.98
IDENTICAL PICTURES	1000.12	130.24	1663.07	1912.20	375.21	286.23	145.54	202.17
CUBES	194.99	140.40	454.51	541.77	360.22	36.10	2.92	11.44
SURFACE DEVELOPMENT	1638.96	643.75	1365.42	689.35	241.98	38.03	241.83	148.23
FORMBOARD	126.44	3.78	323.15	202.87	18.12	1.60	156.27	21.38
PAPER FOLDING	56.05	0.20	348.78	111.13	124.67	50.63	69.50	10.61
OBJECT APERTURE	254.05	3.55	427.30	56.16	284.20	32.40	85.40	21.47
SPATIAL ABILITY	980.63	323.91	6630.58	3159.47	1959.09	133.23	881.73	5.78
MAZES COMPLETED	110.20	3.39	456.97	265.77	538.48	50.63	23.11	335.78
MAZES ERRORS	256.59	0.23	79.30	397.69	10.69	7.23	1437.84	243.81
NUMBER OF PAIRS	66	42	75	68	22	12	47	32
CHI SQUARE	26.88	15.61	52.95	57.37	22.44	17.07	23.63	
PROBABILITY	*	30-50	***	***	†	20-30	†	

TABLE A22: SUMS OF SQUARES DUE TO THE REGRESSION

	MZMW	DZMW	MZFW
ARITHMETIC	19.67	1.28	480.62
VOCABULARY	4.31	2.91	5.39
HEIM	51.23	193.71	29.13
SPELLING	25.21	0.19	4.65
CALENDAR	8.60	11.75	4.47
IDENTICAL PICTURES	81.81	173.38	6.00
CUBES	0.66	84.01	1.28
SURFACE DEVELOPMENT	19.60	17.58	15.93
FORMBOARD	17.53	16.61	5.32
PAPER FOLDING	4.29	9.64	2.34
OBJECT APERTURE	23.86	17.09	16.71
SPATIAL ABILITY	4.00	96.59	7.65
MAZES COMPLETED	42.61	60.11	11.34
MAZES ERRORS	14.40	160.09	42.27
NUMBERS OF PAIRS	66	42	75
CHISQUARE	9.23	16.81	15.07
PROBABILITY	80-90	20-30	30-50

KEY: to column headings

MZ - monozygotic twins
DZ - dizygotic twins
M - males
F - females
W - Whites
B - Blacks

OF WITHIN PAIRS DIFFERENCES IN TEST SCORE ON AGE

DZFW	MZMB	DZMB	MZFB	DZFB
140.74	321.41	561.80	1.45	17.13
4.90	2.38	8.45	6.94	9.49
219.66	4.95	54.45	72.71	132.23
85.66	24.87	2.45	3.90	69.29
59.72	0.69	0.20	26.65	4.62
108.46	60.29	530.45	2.02	472.85
87.54	35.89	12.80	23.34	6.61
12.68	213.54	48.05	32.72	106.33
16.57	12.59	9.80	52.71	8.58
0.35	5.55	4.05	21.84	2.05
37.88	2.16	7.20	0.05	0.62
9.70	6.33	36.45	23.30	9.53
1.68	49.32	8.45	1.74	36.95
73.44	23.58	26.45	160.80	140.46
68.	22	12	47	32.
19.34	23.82	Residual Matrix	15.85	19.11
10-20	•	Singular	30-50	10-20

this matrix, the regression sums of squares, were used to correct the total and between pairs sums of squares from the simple analysis of variance described in Section 2.3.3 as follows:

$$\text{Age corrected between pairs variance} = \frac{\text{Between pairs SS} - \text{Regression SS}}{N - 2}$$

$$\text{Age corrected total variance} = \frac{\text{Total SS} - \text{Regression SS}}{2N - 2}$$

where N is the number of pairs of twins. We notice that one degree of freedom is lost in calculating the regression sum of squares and making the age adjustment. The regression of within pairs variance on age was not significant and no correction was necessary. The corrected total variances and between pairs variances are given in Tables A23 and A24 respectively. These variances and the corresponding uncorrected within pairs variances will be used in all further analysis involving the variances. Thus, the problem of increased variance due to the regression of test score on age has been removed and we will not be overestimating the between families genetical and environmental components of variation because of the age effect.

TABLE A23: AGE CORRECTED TOTAL VARIANCES

	WHITES			
	MALES		FEMALES	
	MZ	DZ	MZ	DZ
NUMBER OF PAIRS	66	42	75	68
ARITHMETIC	614.180	682.699	585.301	603.671
VOCABULARY	11.557	10.822	12.692	9.987
HEIM	351.314	360.776	265.903	258.009
SHEILING	199.380	179.237	128.098	146.072
CALENDAR	53.766	37.788	33.783	36.378
IDENTICAL PICTURES	233.042	226.528	233.703	164.308
CUBES	80.430	86.721	66.113	70.121
SURFACE DEVELOPMENT	186.534	173.479	108.511	107.296
FORMBOARD	56.726	37.951	38.948	40.064
PAPER FOLDING	21.732	21.104	16.293	14.051
OBJECT APERTURE	46.409	51.570	27.003	23.605
SPATIAL ABILITY	335.138	170.522	259.643	330.188
MAZES COMPLETED	68.547	94.921	95.149	65.356
MAZES ERRORS	57.710	85.887	62.165	48.764

BLACKS			
MALES		FEMALES	
MZ	DZ	MZ	DZ
22	12	47	32
851.743	476.136	705.402	857.421
6.582	5.427	8.860	9.153
326.729	328.869	307.955	240.519
234.407	183.658	246.521	207.055
36.296	24.775	45.823	30.341
221.570	237.851	313.715	307.671
59.904	48.291	60.457	25.019
73.569	74.696	41.702	53.028
19.830	18.834	14.012	13.458
16.291	10.063	18.015	13.044
17.761	37.934	14.945	10.895
349.236	197.217	238.523	116.653
53.094	91.669	86.635	81.451
143.384	22.973	233.968	192.556

TABLE A24: BETWEEN AND WITHIN PAIRS VARIANCES (with the between pairs variance corrected for the regression of test score on age)

	WHITES				BLACKS			
	MALES		FEMALES		MALES		FEMALES	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
Arithmetic	1114.6077	997.8075	1059.3918	966.5518	1641.5100	986.6460	1384.4444	1489.6033
	125.9167	382.5952	123.8533	251.4339	133.9127	817.3750	103.1277	264.9000
Vocabulary	18.3423	15.2982	19.9574	11.4064	9.1980	6.6900	12.4246	11.5491
	4.9773	6.5595	5.6200	8.6103	4.2045	4.3750	5.4468	6.9063
Heim	659.6531	523.6025	488.0114	426.2907	599.0550	520.5110	558.2422	381.5900
	52.3182	203.7946	69.7153	94.6765	79.7591	69.1667	68.3192	108.2636
Spelling	382.9359	261.3475	236.3804	237.1833	469.7295	282.2470	441.7686	
	21.3864	101.0357	22.6933	57.6397	20.4778	101.5000	59.5638	
Calendar	85.1878	52.0028	48.0249	51.8044	56.3460	37.0558	72.0598	44.0543
	25.2349	24.2500	19.9200	21.4014	18.0682	14.5417	20.7021	17.4844
Identical Pictures	405.5984	354.8950	467.1164	285.9894	306.4725	245.3250	482.5178	478.8867
	65.7046	104.2938	55.1800	85.6171	144.3864	209.1250	152.0957	47.1563
Cubes	130.1163	132.5665	87.1341	87.3585	78.4935	50.5400	81.7780	27.7557
	32.2500	43.0595	45.6533	53.3597	43.0000	46.4167	40.0436	22.4531
Surface Development	314.5607	275.4200	185.5074	136.6315	115.3200	125.8310	63.2795	65.2903
	33.2955	76.3929	33.5733	78.8235	35.6136	32.0833	21.0436	41.5313
Formboard	93.1163	59.5005	56.7444	53.3259	31.0650	29.6858	13.6696	16.1788
	21.4394	17.4386	21.6267	27.1912	9.6136	9.7917	14.3404	10.9063
Paper Folding	34.4220	30.5000	25.1077	20.5661	26.8119	10.9375	23.4967	4.4417
	9.4242	12.1548	7.7133	7.7279	6.7273	9.3333	7.9787	12.0156
Object Aperture	71.5498	76.5935	40.0136	31.4106	26.8229	59.5055	20.6321	13.6155
	22.0303	27.9286	14.3400	16.0294	9.5221	19.9583	9.5000	8.3138
Spatial Ability	593.3594	615.0325	469.7562	513.3894	670.3700	318.4780	453.7578	174.1987
	84.7424	137.6548	55.1333	152.3750	57.2755	96.1667	32.4468	62.7031
Mazes Completed	96.8375	153.7250	158.4397	86.9812	89.4225	127.0710	121.5418	94.2970
	41.1136	38.9167	33.5467	44.3577	20.0682	62.1667	53.2125	69.4063
Mazes errors	80.8327	98.1065	86.5896	63.3398	232.3865	21.3900	319.8911	233.1017
	35.2879	74.2500	38.1000	34.6177	62.4773	24.2917	151.7021	154.5469

5 ANALYSIS OF THE MEANS

5.1 The Method

We have shown, during the analysis of covariance, that there are significant mean differences between the eight groups of twins even after the means have been adjusted for the effects of age. We now turn to the problem of determining the source of these mean differences. *Differences between the two races, the two sexes* or the two zygosities are immediately obvious sources of overall mean differences. We would also be interested in any interactions between these primary sources of mean differences. For example, we see that White monozygotic males score more than White monozygotic females on the Arithmetic Test. However, the position is reversed in the Blacks. If significant, this race x sex interaction would be of some importance in our study of variation in different racial groups and in the two sexes. A three-way analysis of variance could be used to analyse the means. However, the unequal numbers in the different groups makes this analysis complex and tedious and we shall adopt another approach here, which is formally equivalent to the three-way analysis of variance. We may specify a simple model for the mean differences and then fit this model using weighted least squares. The concepts and methods of weighted least squares are described in Appendix E.

The model assumes coefficients of 1 for the effects since we have no expectations for these coefficients based on theory. The effects are contrasted by specifying the two groups as +1 and -1. For example, for the races effect Whites are specified as +1 and Blacks as -1. The

coefficients of the interactions are obtained by multiplying the coefficients of the appropriate main effects together. The full model is presented in Table A25.

The different groups contain different numbers of subjects and, therefore, the different groups will be weighted in the analysis. The appropriate weights are the inverses of the variances of the means i.e. $\frac{N}{V}$ where N is the number of individuals on which the mean was based and V is the variance of the group. The model was fitted using weighted least squares and estimates of the effects were obtained. These estimates are shown in Table A26. Dividing each estimate by its standard error yields a value which may be used to test the significance of the effects. The standard errors are obtained as the square roots of the diagonal elements of the variance-covariance matrix which is derived from the weight and model matrices. The significance levels of the estimates are given in Table A26. An example of a typical variance-covariance matrix is given in Table A27. The diagonal terms are the variances of the estimates and the off-diagonal terms are the covariances between the estimates. The covariances between the main effects and the covariances between the interactions are small, although the covariance between main effects and interactions involving that main effect are somewhat higher. This is expected since the main effect and the interaction involving the main effect are not independent. when the numbers are unequal. However, apart from this, the race, sex and zygosity effects are independent of one another. This is not always the case in model fitting, but is found here since the rows of the model were specified as orthogonal. It turns out that all the diagonal elements of the matrix are the same and, therefore, the standard errors of all the estimates are identical.

TABLE A25: SIMPLE LINEAR MODEL FOR TESTING DIFFERENCES AMONG THE MEANS

	WHITES				BLACKS			
	Males		Females		Males		Females	
	MZ	DZ	MZ	DZ	MZ	DZ	MZ	DZ
Mean	1	1	1	1	1	1	1	1
Zygosity	1	-1	1	-1	1	-1	1	-1
Race	1	1	1	1	-1	-1	-1	-1
Sex	1	1	-1	-1	1	1	-1	-1
ZxR	1	-1	1	-1	-1	1	-1	1
ZxS	1	-1	-1	1	1	-1	-1	1
RxS	1	1	-1	-1	-1	-1	1	1
ZxRxS	1	-1	-1	1	-1	1	1	-1

TABLE A26: EFFECTS CONTRIBUTING TO DIFFERENCES

	MEAN	ZYGOSITY	RACE
ARITHMETIC	54.989***	1.986NS	11.968***
VOCABULARY	3.808***	0.375**	1.220***
HEIM	32.389***	1.859*	7.823***
SPELLING	28.818***	1.387*	6.117***
CALENDAR	9.989***	0.740**	3.164***
IDENTICAL PICTURES	53.076***	1.213NS	2.884***
CUBES	5.018***	0.079NS	3.122***
SURFACE DEVELOPMENT	20.221***	1.026*	4.761***
FORM BOARD	9.072***	0.232NS	3.064***
PAPER FOLDING	5.598***	0.207NS	2.329***
OBJECT APERTURE	2.883***	0.451NS	2.137***
SPATIAL ABILITY	46.004***	1.121NS	10.337***
MAZES COMPLETED	26.000***	0.530NS	3.317***
MAZES ERRORS	0.826***	0.006NS	-0.033NS

KEY: * significant at the 5% level
** significant at the 1% level
*** significant at the 0.1% level

IN MEANS BETWEEN THE GROUPS OF TWINS

SEX	ZxR	ZxS	RxS	ZxRxS	s.e.
-3.111**	-0.212NS	2.079NS	3.158**	-0.516NS	1.139
0.171NS	0.053NS	0.320*	-0.065NS	-0.056NS	0.133
0.194NS	0.650NS	0.849NS	0.529NS	-1.006NS	0.808
-2.989***	-0.338NS	1.002NS	0.486NS	-0.250NS	0.624
0.007NS	0.022NS	0.597*	0.049NS	0.075NS	0.259
-1.181NS	-0.468NS	-0.358NS	1.554*	0.576NS	0.693
1.192***	0.346NS	-0.095NS	-0.219NS	0.580NS	0.335
0.287NS	0.498NS	0.261NS	0.501NS	0.377NS	0.421
0.764***	0.484*	0.189NS	0.177NS	0.551NS	0.218
0.387*	0.087NS	-0.109NS	-0.117NS	0.239NS	0.175
1.341***	-0.039NS	0.056NS	0.339NS	0.089NS	0.244
3.783***	-1.101NS	0.195NS	0.188NS	0.155NS	0.717
1.034**	0.024NS	-0.063NS	0.440NS	-0.167NS	0.400
-0.065***	-0.036NS	-0.008NS	0.052**	0.003NS	0.019

TABLE A27: VARIANCE-COVARIANCE MATRIX FOR THE ESTIMATES OF THE VOCABULARY TEST

	Mean	Zygosity	Race	Sex	ZxR	ZxS	RxS	ZxRxS
Mean	176	-39	-44	41	27	-27	-27	13
Zygosity	-39	176	27	-27	-44	41	13	-27
Race	-44	27	176	-27	-39	13	41	-27
Sex	41	-27	-27	176	13	-39	-44	27
ZxR	27	-44	-39	13	176	-27	-27	41
ZxS	-27	41	13	-39	-27	176	27	-44
RxS	-27	13	41	-44	-27	27	176	-39
ZxRxS	13	-27	-27	27	41	-44	-39	176

N.B. All figures must be multiplied by 10^{-4} to obtain the actual values

5.2 The Races Effect

Turning to the results presented in Table A26, we see that there is a significant positive race effect for every test. White means are always significantly greater than Black means. Many other studies (e.g. Vandenberg, 1969; Scarr-Salapatek, 1971; Jensen, 1973b) have found significant mean differences between Whites and Blacks in a similar direction. Several workers have suggested that the difference in mean is due to genetical differences between the two populations since the two groups do not intermarry freely and may have been subject to different selective pressures in the past, leading to a different genetic composition (Jensen, 1969, 1973b).

Other workers have pointed to the underpriveleged position of Blacks in American society and suggested that the lower mean of Blacks may be explained purely in terms of environmental deprivation (Lewontin, 1970; Kamin, 1974; Tizard, 1974, 1975). We will not be able to discriminate between these alternative hypotheses about the causation of the lower mean scores of Blacks on tests of ability. However, we shall fit models to the between and within pairs variances of Whites and Blacks and attempt to estimate genetical and environmental components of variation. Thus, although we can say nothing about the causes of differences in means between the races, we may be able to throw some light on the causes of individual differences within the two populations.

5.3 The sexes effect

The mean differences between the eight groups cannot be explained solely in terms of differences between Whites and Blacks. There are significant sex effects for most of the tests. Looking firstly at the Verbal-Arithmetic factor as defined in Section 2.2, there are significant sex effects for only two of the tests: Arithmetic and Spelling. The non-significant sex effect for the other two tests is positive, but the significant sex effects for Arithmetic and Spelling are both negative (i.e. females have a higher mean than males on these tests). It is well known that females score more highly on Verbal tests than males (McNemar, 1942; Anastasi, 1949; Levine et al, 1967; Heim, 1970). However, the finding that females score higher than males on the Arithmetic Test is in conflict with other studies (McNemar, 1942; Terman et al, 1946). We do not have to look far to see why these results are in conflict with previous studies. Previous studies used White subjects only. For the Arithmetic test we have a highly significant races x sexes interaction in these data. In other words, males score more highly than females in the White group, as found in previous studies, but females score more highly than males in the Black group. Jensen (1972, 1973b) has reviewed studies which show a races x sexes interaction for measures of ability (Brofenbrenner, 1967). Since the difference between the means of the two sexes is higher in the Black group than in the White group, the sex effect itself turns out to be negative overall.

Turning now to the tests of the Spatial factor, we see that seven of the ten tests loading on this factor show a significant positive sex effect i.e. males have a higher mean than females in all cases.

Many workers have shown that males do "better" in Spatial tasks than females and the cause of the sex difference has been a matter of some controversy. Some workers have argued that females score less highly because they have less educational opportunities than males. This seems unlikely since Spatial tests load very highly on Jensen's Level II ability and not so highly on Level I ability, In other words, learning improves performance on this type of test only very slightly. Thus, most workers look for an explanation of this effect in some genetic mechanism. The difference in mean could be due to sex-limitation or sex-linkage. Many workers have argued that enhancement of spatial ability is due largely to a single recessive sex-linked gene. (Stafford, 1961; Hartlage, 1970; Bock and Kolakowski, 1973).

If this were true then females would score lower than males on average because the frequency of males showing the enhancement effect would equal the frequency of the gene in the population, but the frequency of the females showing the enhancement effect would be the square root of the gene frequency. Bock and Kolakowski (1973) have presented evidence for sex-linkage, but this is by no means conclusive. The most usual tests for sex-linkage come from data on parents and offspring. However, there is a test for sex-linkage in these data, albeit not a very powerful one, and tests for sex-linkage will be discussed at length in Section 3.

5.4 The Zygosities Effect

Monozygotic and dizygotic twins are drawn from the same population and there are no a priori reasons to suspect that they should have

different means. However, from Table A24, we see that monozygotic twins score significantly more, on average, than dizygotic twins for five of the fourteen tests. Four of these five tests load on the Verbal-Arithmetic factor. We recall that Husén (1969) tried to show that the lower mean score of twins compared with singletons was due to the lower score of monozygotic twins because of their reduced communication with the outside world. In fact, he showed the reverse. He found that the mean score of monozygotic twins was higher than that of dizygotic twins on verbal tests, although he advanced no reason for this.

Bulmer (1970) summaries evidence that dizygotic twins tend to be born to mothers from the lower socio-economic status groups, whilst the monozygotic twinning rate is the same in all socio-economic status groups. Since the mean score of lower socio-economic status groups in Verbal ability is lower than that of higher groups then,all other things being equal, dizygotic twins will have a lower mean than monozygotic twins.

In this analysis of the complete data of the Georgia Twin Study, we cannot determine the cause of the higher mean score of monozygotic twins. However, in the Extended Study, data is available on the socio-economic status of the twins' parents, and thus the possibility that the lower mean score of dizygotic twins is due to the lower socio-economic status of their parents could be investigated. This investigation will be carried out at a later date.

5.5 Other Interactions

Five interactions between main effects are significant at the five

per cent level. Since we would expect several interactions out of the possible fifty-five interactions to be significant at the five per cent level by chance, we cannot attach any importance to these interactions. They do not fall into any consistent pattern and since we had no prior expectation for any particular interaction, it would be impossible to attach any confidence to post hoc rationalisations of these interactions.

SECTION 3: MODELS AND MODEL FITTING

1 INTRODUCTION

We have briefly described the means, variances and coefficients of skewness and kurtosis found in these data. These are known respectively as first, second, third and fourth degree statistics and, given information on theappropriate groups of individuals and a sufficient theoretical framework, expectations for these statistics in terms of meaningful genetical and environmental effects can be formulated and these effects can be estimated.

Expectations for the means of inbred lines and families of known genotypes derived from them can be formulated and hence genetical and environmental effects can be estimated. However, models for the means of natural populations cannot be formulated since we do not know the genotypes of the individuals in the population. Therefore, expectations for the contribution of genetical and environmental effects to the covariance between different types of relatives are specified and estimated instead. Expectations in terms of genetical and environmental effects can also be specified for the third and fourth moments about the mean (i.e. skewness and kurtosis). However, as higher moments are used, the standard errors of estimates obtained increase disproportionately. Therefore, more precise estimates can be obtained using means than any other statistic. However, in the case of twin studies only variances can be used.

In formulating models for the contribution of genetical and environmental components of variation to the covariation between

monozygotic and dizygotic twins, we shall follow the approach and notation initiated by Mather (Mather, 1949) and later extended by Mather and Jinks (Mather and Jinks, 1971). Firstly we shall specify an empirical model which may be used, for simplicity, in preliminary analysis of data. This model is discussed in some detail by Jinks and Fulker (1970), and is limited in its application since it is unable to tell us anything about gene effects, mating systems or past selection and evolution of the trait. Then we shall consider the model of Mather and Jinks which allows us to specify parameters for additive gene effects, dominance, epistasis, genotype-environment interactions and covariation, epistasis and assortative mating. Therefore, by fitting the appropriate model we can determine the genetic architecture of the trait in question and make inferences about the direction of selection upon the trait in the past and about its evolution (Mather, 1943).

2 THE SIMPLE EMPIRICAL MODEL

2.1 Parameters of the Simple Empirical Model

A simple empirical model can be specified and its parameters estimated providing the assumptions on which it is based are recognised and tested. In general, four components of variation may be recognised. We may distinguish between genetical influences producing variation between families, called G_2, and those producing differences within families, called G_1. The G_2 effect arises because each family shares a common pool of parental genes, different to those of every other family. This produces covariation between members of a family and produces genetical differences between families since each family has a different set of parental genes. The processes of independent assortment of chromosomes and recombination during meiosis ensure that the members of a family each receive a different set of the parental genes. There is, thus, a source of variation within families (G_1).

Similarly, there are environmental influences specific to individuals which produce variation within families (E_1), and those common to families as a whole (E_2). Examples of factors that might contribute to E_2 are socio-economic status of home, educational level of parents etc. Factors contributing to E_1 are accidents of development, both pre-natal and post-natal, test unreliability and the experiences of one sib not shared by the other.

2.2 Contribution of these Parameters to Different Statistics

The total variance may be defined in terms of the four parameters specified above as follows: $\sigma_T^2 = G_1 + G_2 + E_1 + E_2$. We now turn to the problem of the relative contribution of these parameters to different types of family. We shall consider five types of family commonly used in research into human behaviour, although we have data on only two types of family in this study. These are monozygotic twins reared together and apart, dizygotic twins (or full-sibs) reared together and apart and unrelated individuals reared together.

The total variances of the five types of family are expected to be equal unless the assumptions underlying this simple model fail. These assumptions will be discussed later. We can formulate precise expectations for the partition of the total variances into between pairs variance (σ_b^2) and within pairs variance (σ_w^2) for the five family types.

The effects of E_1 are specific to individuals and always contribute to variation within pairs. The common environmental effect, or E_2, is due to environmental influences common to individuals reared together in the same home. Therefore, E_2 contributes to variation between pairs of individuals when they are reared together, but contributes to variation within pairs when the members of the pair are raised in different homes.

The G_2 effect arises because the offspring of each pair of parents share a common pool of parental genes. Therefore, G_2 always contributes to variation between pairs of twins. However, unrelated individuals reared together do not share a common pool of parental genes and G_2

contributes to σ_w^2 in this case. Monozygotic twins have an identical selection from the parental genes, therefore, G_1 contributes to variation between pairs for monozygotic twins. Dizygotic twins have different combinations of parental genes and G_1 contributes to σ_w^2. The G_1 effect also contributes to within pair differences for unrelated individuals reared together since they have different sets of genes. These expectations for σ_b^2 and σ_w^2 which we have described verbally are summarised in Table A28a.

In our analysis of variance in Section 2.3.3 we calculated between pairs and within pairs mean squares. The expected mean squares in terms of the variance components are given:

Source	d.f.	Expected Mean Square
Between pairs	n-1	$\sigma_w^2 + 2\sigma_b^2$
Within pairs	n	σ_w^2

where n is the number of pairs. We can now formulate our expectations for the between pairs and within pairs mean squares of the different family types from our expectations for the σ^2's. The expectations for these mean squares are given in Table A28b.

If we had data on these five types of family, we could obviously estimate all four parameters of our model. In fact, using these ten statistics we could obtain several estimates of each parameter. Whenever we do not have what is known as a "perfect fit solution", (where the number of statistics equals the number of parameters that can be estimated), then a procedure known as weighted least squares is used to obtain the "best", or maximum-likelihood estimates, of the parameters. This procedure which is discussed in detail later, allows

TABLE A28a: EXPECTATIONS FOR THE σ^2's OF FIVE TYPES OF FAMILY

FAMILY TYPE	σ^2	G_1	G_2	E_1	E_2
Monozygotic Twins reared together (MZ_T)	σ^2_b	1	1	-	1
	σ^2_w	-	-	1	-
Monozygotic Twins reared apart (MZ_A)	σ^2_b	1	1	-	-
	σ^2_w	-	-	1	1
Dizygotic Twins reared together (DZ_T)	σ^2_b	-	1	-	1
	σ^2_w	1	-	1	-
Dizygotic Twins reared apart (DZ_A)	σ^2_b	-	1	-	-
	σ^2_w	1	-	1	1
Unrelated Individuals reared together (U_T)	σ^2_b	-	-	-	1
	σ^2_w	1	1	1	-

TABLE A28b: EXPECTATIONS FOR THE MEAN SQUARES OF FIVE TYPES OF FAMILY

Family Type	Mean Square	G_1	G_2	E_1	E_2
MZ_T	Between pairs	2	2	1	2
	Within pairs	-	-	1	-
MZ_A	Between pairs	2	2	1	1
	Within pairs	-	-	1	1
DZ_T	Between pairs	1	2	1	2
	Within pairs	1	-	1	-
DZ_A	Between pairs	1	2	1	1
	Within pairs	1	-	1	1
U_T	Between pairs	1	1	1	2
	Within pairs	1	1	1	-

us to estimate the parameters of the model and their standard errors so that we can test various hypotheses about the estimates (e.g. whether they are significantly greater than zero). We can also test the adequacy of the model, when the number of parameters is less than the number of statistics. This is one of the greatest strengths of the biometrical-genetical approach.

There are several minimal sets of data which allow us to estimate the four parameters and test the adequacy of the model. For example we could use the six statistics provided by MZ_T, MZ_A and DZ_T, or alternatively the six statistics provided by MZ_T, DZ_T and DZ_A. Both these sets of data enable us to estimate G_1, G_2, E_1 and E_2 and to test the adequacy of the model.

2.3 Parameters that can be Estimated with Data of this Study

In this study we have only two types of family - monozygotic twins reared together and dizygotic twins reared together. Therefore, we have four statistics for estimating the parameters of the model and testing the adequacy of the model. From Table A28a, we see that we can estimate E_1 as ($\hat{\sigma}_w^2$ of MZ_T) and G_1 as ($\hat{\sigma}_w^2$ of DZ_T minus $\hat{\sigma}_w^2$ of MZ_T). However, the coefficients of G_2 and E_2 are the same for all statistics and we cannot estimate these effects separately. The best we can do is to estimate $(G_2 + E_2)$, which we shall consider to be an estimate of G_2 biased by any E_2 effects. We shall call this biased estimate B i.e. $\hat{B} = G_2 + E_2$. $\hat{B}$ is also biased by any covariation or interaction between G_2 and E_2, as we shall show later. We can now obtain estimates of our three parameters, G_1, B and E_1 and there is one degree of freedom left for testing the adequacy of the model.

The broad heritability of a trait is defined as that proportion of the variation for a trait due to genetical causes. This is given by:

$$\hat{h}_b^2 = \frac{G_1 + G_2}{G_1 + G_2 + E_1 + E_2}$$

In this study, we cannot estimate h_b^2, but only a biased version of it:

$$\hat{Bh}_b^2 = \frac{G_1 + B}{G_1 + B + E_1}$$

The quantity $\hat{Bh}_b^2$ is the upper limit for h_b^2 and equals h_b^2 only where E_2 equals zero. If E_2 is not zero, then $\hat{Bh}_b^2$ is h_b^2 inflated by common environmental effects.

The effects of G_2 and E_2 can only be separated in studies where data on relatives reared apart is available. Several such studies of general intelligence, where there was data on monozygotic twins reared apart (e.g. Burt, 1966; Shields 1962; Newman, Freeman and Holzinger, 1937) have suggested that the E_2 effect is small, accounting for less than five per cent of the variation. If this estimate of E_2 is about right for the tests of ability used in this study, then $\hat{Bh}_b^2$, our estimate of h_b^2, will not be grossly inflated by E_2. However, these studies have been severely criticised on a number of grounds, including the type of tests used, poor separation of the monozygotic twins and the effects of age, for which no correction was made. Therefore, we must exercise caution in generalising the finding of a negligible $\hat{E}_2$ in these previous studies to our study, especially since we have tests of Verbal-Arithmetic ability and Spatial ability rather than of general intelligence. We must recognise that our estimate of h_b^2, $\hat{Bh}_b^2$, is a biased estimate inflated by common environmental effects.

2.4 Assumptions underlying the Simple Model and Tests for them

The simple empirical model which we have described and discussed embodies four basic assumptions. These are enumerated by Jinks and Fulker (1970):

1. No genotype-environment interaction
2. No genotype-environment covariation
3. The total genetic variance, $G = G_1 + G_2$
4. The E_1's and E_2's are the same for all kinds of families.

Jinks and Fulker give tests for GE_1 interactions and GE_2 interactions which involve looking for a relationship between absolute pair differences and pair sums of monozygotic twins reared together for GE_1 interactions and monozygotic twins reared apart for GE_2 interactions. These tests will be described in depth in Section 6, and we shall just note here that such tests for systematic genotype-environment interaction can be made and failure of the assumption of no systematic genotype-environment interaction can be detected.

The last three assumptions can be tested jointly since they each lead to heterogeneity of the total variances and probably arise from the same cause. The total variances may be tested directly for heterogeneity using Bartlett's or some similar test (Winer, 1962); an analysis of the variance between different types of family may be carried out; or a test of the "goodness of fit" of the estimated parameters may be made during the model fitting procedure (see Appendix E). If heterogeneity of the total variances is detected one of the assumptions described above has failed. Eaves (1976a, b) has shown that genotype-environmental covariation due to sibling effects (either competition or

cooperation) or due to cultural transmission each lead to a predictable pattern of inequality of the total variances. Thus, failure of assumption 2 can be detected. Genotype-environment covariation due to sibling effects may produce inequality of the E_1's and E_2's between different groups, and hence also lead to the failure of assumption 4. If inequality of the total variances is not due to failure of the assumption of no genotype-environment covariation, then either $G \neq G_1 + G_2$ or else the E_1's and E_2's differ between groups. A test for determining which of these assumptions has failed has not been proposed and would be difficult in practise. Either E_1 or E_2 or both could differ with family type, and could take a different value for every type of family. We have no theoretical expectations for the relative magnitudes of different environmental components in different families, except in the case of sibling effects, and, therefore, we cannot predict the pattern of inequality of the total variances that would be produced by failure of the assumption that E_1's and E_2's are equal for all families, unless the failure is due to sibling effects. Similarly the inequalities among the total variances that would be produced by failure of assumption 3 cannot be predicted theoretically. Therefore, if we show that failure of the assumption of homogeneity of the total variances is not produced by genotype-environment covariation, the only course of action is to specify empirical models and attempt to fit these to the data. This procedure is not very satisfactory since it is not based on any theoretical framework, and is only a post hoc rationalisation of a body of data which will not enable us to make predictions about further types of relatives or future studies. Urbach (1974a,b) describes a science which depends on

such post hoc rationalisation as degenerate and says that "progress" can only be made where a theoretical framework enables precise expectations to be formulated and tested.

2.5 Specification of Genotype-Environment Interaction and Genotype-Environment Covariation in the Simple Empirical Model

Genotype-Environment Interaction

Genotype-environment interaction arises when the effect of E_1 or E_2 or both is under genetical control. We may recognise four interactions: $G_1 \times E_1$, $G_2 \times E_1$, $G_1 \times E_2$ and $G_2 \times E_2$. The first two types of interaction may, if the interaction is systematic, lead to a relationship between pair sums and pair differences for monozygotic twins reared together and the last two types will lead to a similar relationship for monozygotic twins reared apart (Jinks and Fulker 1970). Given that we have detected genotype-environment interaction, we may specify its contribution to the σ's and mean squares of different types of relatives. These expectations are shown in Table A29. The first point to notice is that genotype-environment interactions alter the relative magnitude of the σ_b^2's and σ_w^2's but do not lead to inequality of the total variances. Secondly, in order to estimate all the parameters of this extended model and to test the adequacy of this model, at least nine statistics would be needed. (However, all eight parameters cannot be estimated since E_1 and G_1 are confounded in the presence of family groupings. Therefore, in practise, only eight statistics would be needed).

TABLE A29: SPECIFICATION OF GENOTYPE-ENVIRONMENT INTERACTIONS

EXPECTATIONS FOR THE σ^2's

Type of Family	σ^2	G_1	G_2	E_1	E_2
MZ_T	σ^2_b	1	1	-	1
	σ^2_w	-	-	1	-
MZ_A	σ^2_b	1	1	-	-
	σ^2_w	-	-	1	1
DZ_T	σ^2_b	-	1	-	1
	σ^2_w	1	-	1	-
DZ_A	σ^2_b	-	1	-	-
	σ^2_w	1	-	1	1
U_T	σ^2_b	-	-	-	1
	σ^2_w	1	1	1	-

EXPECTATIONS FOR THE MEAN SQUARES

Type of Family	Mean Square	G_1	G_2	E_1	E_2
MZ_T	Between pairs	2	2	1	2
	Within pairs	-	-	1	-
MZ_A	Between pairs	2	2	1	1
	Within pairs	-	-	1	1
DZ_T	Between pairs	1	2	1	2
	Within pairs	1	-	1	-
DZ_A	Between pairs	1	2	1	1
	Within pairs	1	-	1	1
U_T	Between pairs	1	1	1	2
	Within pairs	1	1	1	-

$G_1 \times E_1$	$G_1 \times E_2$	$G_2 \times E_1$	$G_2 \times E_2$
–	1	–	1
1	–	1	–
–	–	–	–
1	1	1	1
–	–	–	1
1	1	1	–
–	–	–	–
1	1	1	1
–	–	–	–
1	1	1	1

$G_1 \times E_1$	$G_1 \times E_2$	$G_2 \times E_1$	$G_2 \times E_2$
1	2	1	2
1	–	1	–
1	1	1	1
1	1	1	1
1	1	1	2
1	1	1	–
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1

We can obtain an estimate of G_2 from (σ^2_b for DZ_A) and an estimate of E_2 from (σ^2_b for U_T). An estimate of G_1 may be obtained as the difference between (σ^2_b for MZ_A) and (σ^2_b for DZ_A). Hence we can estimate $G_2 \times E_2$ from (σ^2_b for DZ_T) and $G_1 \times E_2$ from MZ_T. We notice that we are unable to separate the effects of E_1, $G_1 \times E_1$ and $G_2 \times E_1$ with these five types of family and, indeed, it seems that the effects of E_1 are inevitably confounded with those of $G_1 \times E_1$ and $G_2 \times E_1$ in the presence of family groupings (Mather and Morley-Jones, 1958).

Genotype-environment covariation

Genotype-environment covariation may arise when "good" genotypes tend to obtain "good" environments and "poor" genotypes tend to obtain "poor" environments. This type of covariation increases the differences between genotypes in terms of their phenotypic expression. Genotype-environment covariation may also be of the type where "good" genotypes have "poor" environments and "poor" genotypes of "good" environments. This type of covariation makes individuals phenotypically more similar. We may recognise four possible covariance terms between genotype and environment: $COV(G_1E_1)$, $COV(G_1E_2)$, $COV(G_2E_1)$ and $COV(G_2E_2)$. If we detect genotype-environment covariation in a body of data, we could specify the contribution of these four parameters to the σ^2's and hence to the mean squares and attempt to estimate these parameters. This purely statistical approach to the problem of genotype-environment covariation is that adopted by Cattell (1960, 1965b) and is inadequate since it is purely empirical. Eaves (1976a, b) has proposed a theory

of the mechanism of genotype-environment covariation which may lead to a more satisfactory analysis of g-e covariation. The expectations of the empirical model for data showing genotype-environment covariance are given in Table A30.

We see that the contribution of the genotype-environment covariance parameters to the total variances are not the same and, therefore, genotype-environment covariation leads to inequality of the total variances. We may estimate all the parameters of this model and test its adequacy given nine appropriate statistics. However, we must remember that certain types of genotype-environment covariation lead to inequality of the E_1's and E_2's between groups. In these cases, the model of Eaves (1976a,b) is more appropriate.

We see that E_1, E_2 and G_2 may be estimated directly as (σ^2_w of MZ_T), (σ^2_b of U_T) and (σ^2_b of DZA) respectively. Then G_1 may be obtained from (σ^2_b of MZA). Now $COV(G_2E_2)$ and $COV(G_1E_1)$ may be estimated from (σ^2_b and σ^2_w for DZ_T), leaving only $COV(G_1E_2)$ to be estimated from (σ^2_b for MZ_T) and $COV(G_2E_1)$ to be estimated from (σ^2_w of U_T). Thus, we do not have the problem, found in the case of genotype-environment interaction that certain parameters can never be estimated in the presence of family groupings.

2.6 Fitting models to monozygotic and dizygotic twins reared together

The effect of Genotype-Environment Interaction

In this study, we have only monozygotic and dizygotic twins reared together. These two families provide four statistics only and, therefore,

TABLE A30: SPECIFICATION OF GENOTYPE-ENVIRONMENT COVARIATION

TYPE OF FAMILY	σ^2	G_1	G_2	E_1	E_2
MZ_T	σ^2_b	1	1	-	1
	σ^2_w	-	-	1	-
MZ_A	σ^2_b	1	1	-	-
	σ^2_w	-	-	1	1
DZ_T	σ^2_b	-	1	-	1
	σ^2_w	1	-	1	-
DZ_A	σ^2_b	-	1	-	-
	σ^2_w	1	-	1	1
U_T	σ^2_b	-	-	-	1
	σ^2_w	1	1	1	-

EXPECTATIONS FOR THE MEAN SQUARES

Type of Family	Mean Square	G_1	G_2	E_1	E_2
MZ_T	Between pairs	2	2	1	2
	Within pairs	-	-	1	-
MZ_A	Between pairs	2	2	1	1
	Within pairs	-	-	1	1
DZ_T	Between pairs	1	2	1	2
	Within pairs	1	-	1	-
DZ_A	Between pairs	1	2	1	1
	Within pairs	1	-	1	1
U_T	Between pairs	1	1	1	2
	Within pairs	1	1	1	-

$COV(G_1E_1)$	$COV(G_1E_2)$	$COV(G_2E_1)$	$COV(G_2E_2)$
-	1	-	1
-	-	-	-
-	-	-	-
-	-	-	-
-	-	-	1
1	-	-	-
-	-	-	-
1	1	-	-
-	-	-	-
1	-	1	-

$COV(G_1E_1)$	$COV(G_1E_2)$	$COV(G_2E_1)$	$COV(G_2E_2)$
-	2	-	2
-	-	-	-
-	-	-	-
-	-	-	-
1	-	-	2
1	-	-	-
1	1	-	-
1	1	-	-
1	-	1	-
1	-	1	-

we cannot estimate the parameters of the complex model involving genotype-environment interaction. It turns out that we can estimate the following three quantities by fitting our G_1, B, E_1 model to the data.

1. $(E_1 + G_1E_1 + G_2E_1) = \hat{E}_1$

2. $(G_1 + G_1 \times E_2) = \hat{G}_1$

3. $(G_2 + E_2 + G_2 \times E_2) = (B + G_2 \times E_2) = \hat{B}$

Thus in the presence of genotype-environment interaction, we are unable to estimate any of the interactions and our estimates of $\hat{E}_1$, $\hat{G}_1$ and $\hat{B}$ are biased by the interactions, as described above. Our biased estimate of the broad heritability $\hat{Bh}^2_b$ now becomes:

$$\hat{Bh}^2_b = \frac{(G_1 + G_1 \times E_2) + (B + G_2 \times E_2)}{(G_1 + G_1 \times E_2) + (B + G_2 \times E_2) + (E_1 + G_1E_1 + G_2E_1)}$$

We notice that this estimate is now further inflated by interactions between the genotype and E_2.

The effect of Genotype-Environment Covariation

If the assumption of no genotype-environment covariation fails and we fit our three parameter empirical model to the data, we obtain estimates of the following three quantities:

1. E_1

2. $G_1 + COV(G_1E_1)$

$$3.\quad (G_2 + E_2 + COV(G_2E_2)) = (B + COV(G_2E_2)) = \hat{B}$$

Thus, the only estimate which is not biased by the presence of genotype environment covariation is E_1. Therefore, Bh^2_b becomes:

$$Bh^2_b = \frac{G_1 + COV(G_1E_1) + B + COV(G_2E_2)}{G_1 + COV(G_1E_1) + B + COV(G_2E_2) + E_1}$$

Genotype-environment covariation, therefore, inflates our biased estimate of the broad heritability still further. However, since genotype-environment interaction and genotype-environment covariation may be regarded as genetical effects, this may not be a serious problem.

3 THE THEORETICAL MODEL

The simple, empirical model described in Section 3.2 is useful during preliminary analysis. However, its use is limited since it enables us to say nothing about gene action and interaction or the mating system for the traits we are interested in. Therefore, we can say nothing about past selection on the trait or its evolution. It also allows us to say very little about the sources of differences in means and variances between groups, such as the Whites and Blacks of this study. Therefore, we shall define the parameters of the theoretical, biometrical-genetical model of Mather and Jinks (1971) and show how the parameters of the two models are related.

We consider first the simplest genetical model where all the gene action is additive and mating is random. The contribution of the additive effects of all the genes producing variation in the trait we are interested in to the total variance is given by $\frac{1}{2}D_R$,

where $\frac{1}{2}D_R = \Sigma 4u_i v_i d^2_i$

and u_i is the frequency of the increasing allele at the ith locus affecting the trait.

v_i is the frequency of the decreasing allele at the ith locus and $u_i + v_i = 1$

d_i is the additive deviation of the ith locus i.e. the differences between its two homozygotes.

In this case, the total genetic variance, G, may be specified as follows:

$$G = G_1 + G_2 = \tfrac{1}{2}D_R$$

and $G_1 = G_2 = \tfrac{1}{4}D_R$. The broad heritability is given by:

$$\hat{h}^2_b = \frac{\tfrac{1}{2}D_R}{\tfrac{1}{2}D_R + E_1 + E_2}$$

The narrow heritability is the proportion of the total variation due to the additive genetical effects and equals the broad heritability for this simple model.

If this simple model is not adequate to explain variation in the trait we are interested in, then we might suspect dominance for the trait. Now, allowing for dominance, which is the interaction between two alleles at a locus, we may define the contribution of genetical variation to the total variation:

$$G = G_1 + G_2 = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R$$

$$= \Sigma 2u_i v_i [d_i + (v_i - u_i)h_i]^2 + \Sigma 4u_i v_i h_i^2$$

where h_i is the dominance deviation of the ith locus. But now $G_1 \neq G_2$. It can be shown that

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

and $$G_2 = \tfrac{1}{4}D_R - \frac{1}{16} H_R$$

Thus, from the theoretical model, we can now begin generating further expectations for the empirical model. If there is no dominance we expect $G_1 = G_2$. However, in the presence of dominance we expect $G_1 > G_2$ by $1/8H_R$. Therefore, we have a test for dominance which can be

applied to our estimates from the empirical model. Eaves (1972) has shown that H_R and E_2 are inseparable in twins reared together. So, the broad heritability defined as $\hat{h}^2_b = \frac{\frac{1}{2}D_R + \frac{1}{4}H_R}{\frac{1}{2}D_R + \frac{1}{4}H_R + E_1 + E_2}$ is biased in the presence of E_2. The narrow heritability is given by:

$$\hat{h}^2_n = \frac{\frac{1}{2}D_R}{\frac{1}{2}D_R + \frac{1}{4}H_R + E_1 + E_2}.$$

We can formulate the narrow heritability for the empirical model, if the additive-dominance model holds:

$$\hat{h}^2_n = \frac{3G_2 - G_1}{G_1 + G_2 + E_1 + E_2}$$

The simple theoretical model involving additive and dominance effects can be extended to include the effects of interactions between loci (Fisher, 1918; Kempthorne, 1957; Mather, 1975). However, these are unlikely to account for a large proportion of the total variation and are substantially confounded with D_R and H_R, so we shall consider instead the consequences of non-random mating in the form of assortative mating between spouses, for our model. This is likely to be a major source of variation for measures of ability since estimates of the phenotypic correlation between spouses of between 0.3 and 0.6 have been found for general intelligence (Vandenberg, 1972). However, it will only produce increased additive genetical variance if phenotypic assortative mating produces genotypic assortative mating. Fisher (1918) has shown that:

$$A = h^2_n$$

for the case in which the correlation between phenotypes is a primary factor, where μ is the marital correlation (i.e. the phenotypic correlation between spouses), and A is the correlation between the additive genetical deviations of spouses.

He has specified the contribution of assortative mating, for a polygenic trait, to the total genetic variance, when equilibrium has been reached:

$$G = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R + \tfrac{1}{2}(A/(1-A))\ D_R$$

If mating is random, A=0, and the upper limit of A is given by $A = \mu$, when $h^2_n = 1$

Since assortative mating produces variation between families only, then:

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

and $$G_2 = \tfrac{1}{4}D_R + {}^{1}/16\ H_R + \tfrac{1}{2}\ (A/(1-A))\ D_R$$

We can make further predictions about the relative magnitudes of G_1 and G_2 from the theoretical model. If there is no dominance for our trait, but there is assortative mating, then we expect $G_2 > G_1$ by $\frac{1}{2}A/(1-A))D_R$. Thus, if after estimating G_1 and G_2, we find that they differ significantly, we have evidence of either dominance or assortative mating, depending on which estimate is the greater. If $G_1 = G_2$ we may not conclude that there is no dominance or assortative mating since their effects may cancel if ${}^{1}/8\ H_R \simeq \tfrac{1}{2}\ (A/(1-A))D_R$ (Jinks and Fulker, 1970).

If theoretical genetical models involving additive gene effects, dominance and assortative mating do not adequately explain our data,

we may specify additional parameters specifying non-allelic interactions, genotype-environment interactions or genotype-environment covariation.

We shall consider the relative contributions of the parameters we have specified to the σ's and to the mean squares. Their contributions to the σ's are given in Table A31 and the contributions to the σ's are combined to give the contributions to the mean squares as shown in Table A32. The expectations for the environmental effects are identical to those for the simple empirical model. To see how we determine which of the parameters of this model are necessary to explain variation in our trait, we shall move onto a discussion of model fitting.

TABLE A31: EXPECTATIONS OF THE σ^2's IN TERMS OF GENE EFFECTS

FAMILY TYPE	σ^2	D_R	H_R	$(A/(1-A))D_R$	E_1	E_2
MZ_T	σ^2_b	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	-	1
	σ^2_w	-	-	-	1	-
MZ_A	σ^2_b	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	-	-
	σ^2_w	-	-	-	1	1
DZ_T	σ^2_b	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{2}$	-	1
	σ^2_w	$\frac{1}{4}$	$\frac{3}{16}$	-	1	-
DZ_a	σ^2_b	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{2}$	-	-
	σ^2_w	$\frac{1}{4}$	$\frac{3}{16}$	-	1	1
U_T	σ^2_b	-	-	-	-	1
	σ^2_w	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	-

TABLE A32: EXPECTATIONS FOR THE MEAN SQUARES IN TERMS OF GENE EFFECTS

FAMILY TYPE	MEAN SQUARE	D_R	H_R	$(A/(1-A))D_R$	E_1	E_2
MZ_T	Between pairs	1	$\frac{1}{2}$	1	1	2
	Within pairs	-	-	-	1	-
MZ_A	Between pairs	1	$\frac{1}{2}$	1	1	1
	Within pairs	-	-	-	1	1
DZ_T	Between pairs	$\frac{3}{4}$	$\frac{3}{16}$	1	1	2
	Within pairs	$\frac{1}{4}$	$\frac{3}{16}$	-	1	-
DZ_A	Between pairs	$\frac{3}{4}$	$\frac{3}{16}$	1	1	1
	Within pairs	$\frac{1}{4}$	$\frac{3}{16}$	-	1	1
U_T	Between pairs	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2
	Within pairs	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	-

4 FITTING MODELS BY WEIGHTED LEAST SQUARES

4.1 Model Fitting

We have discussed two simple genotype-environmental models, one empirical and one theoretical, whose parameters may be estimated if we have the appropriate data. If the number of parameters to be estimated equals the number of statistics, there is only one estimate of each parameter and this is the Maximum-Likelihood Estimate. In this case we have a "perfect fit" solution, but no test of the model. However, where we have more statistics than parameters, several different estimates of the parameters are possible and we need a method for determining the "best" estimates of the parameters. Maximum-likelihood estimates of our parameters may be obtained by the method of weighted least squares. The concept of maximum-likelihood estimation and the solution of the problem by the method of weighted least squares are discussed in Appendix E.

Weighted least squares solves the linear equation:

$$\underset{\sim}{X} = \underset{\sim}{A}\,\underset{\sim}{W}\,\underset{\sim}{\theta} + \underset{\sim}{e}$$

where $\underset{\sim}{X}$ is an (nxl) vector of observations

$\underset{\sim}{A}$ is an (nxk) matrix of the coefficients of the model

$\underset{\sim}{W}$ is an (nxn) diagonal matrix of weights

$\underset{\sim}{\theta}$ is a (kxl) vector of parameters of the model

$\underset{\sim}{e}$ is an (nxl) vector of "error" random variables where

$\mathcal{E}(e) = 0$

Usually when we are attempting to estimate the parameters of a model from a set of statistics, each statistic is based on a different number of observations. Therefore, the reliability of the statistics differs and in order to attribute more importance to statistics based on larger numbers of observations, the statistics areweighted by the inverse of their variances. Solution of the linear equation by weighted least squares leads to estimates of the parameters. The standard errors of these estimates may be obtained from the variance-covariance matrix (see Appendix E) and therefore, a c-test can be made to determine whether the parameter estimates differ significantly from zero. Observed statistics are compared with their expectations on the basis of the model by assuming that the sum of squares which we are minimising in the fitting process is distributed as χ^2 with (n-k) degrees of freedom. This chisquare is a measure of the "goodness of fit" of our model. When it is significant the model whose parameters we have estimated is not adequate to explain the observed variation and other more appropriate models must be tested.

Even in the case of a perfect fit solution, where we are unable to test the adequacy of our model, it is useful to fit the model by weighted least squares and to test the significance of the parameters. If any parameters turn out to be non-significant, a reduced model without these parameters can be formulated, fitted and tested for adequacy. If the reduced model fits, we can assume that the original model contained redundant parameters accounting for only a negligible proportion of the total variation.

4.2 Fitting the empirical and the theoretical model

When we fit models to a body of data, we try to explain the variation in the data by the simplest possible model. Therefore, we fit the smallest number of parameters which will adequately explain the variation. We normally begin by attempting to fit E_1 only to the data. The E_1 model embodies the assumption that all variation in a trait is specific to individuals and tests the hypothesis that the between families and within families variances are equal. When this model fails, a model including both E_1 and E_2 effects is fitted. This model assumes there is environmental variation between families. Since this variation is environmental, the model tests the hypothesis that $\sigma_b^2 : \sigma_w^2$ is the same for all types of family. If both these models are inadequate to explain the variation in the data, we conclude that the trait may have a heritable component. Therefore, if we are considering the empirical model, we now fit G_1, G_2, E_1 and E_2. We cannot fit G_1 or G_2 alone, since if there is genetical variation, it will produce variation both between and within families. In this study, where we only have data on monozygotic and dizygotic twins reared together, if the E_1 and E_1E_2 models fail we fit the G_1, B, E_1 model.

In the case of the theoretical model, if the E_1 and E_1E_2 model fails, we have an alternative two parameter model that we can fit. This is the simple additive-genetical model - the E_1D_R model. If the variation cannot be explained in terms of E_1 and D_R alone, we attempt to fit three parameter models. Several three parameter models are possible. These are the $E_1E_2D_R$ model, the $E_1D_RH_R$ model and the E_1D_RA model. If these models do not fit, we specify additional

parameters and test the new models until we find a model that fits the data. It may turn out that parameters for epistasis, genotype-environment interaction or genotype-environment covariation are necessary to explain variation in the trait.

One problem remains to be discussed. It is possible that we may find two models with an equal number of parameters that are adequate to explain variation in the trait. How do we decide between the two alternative models? For example, the $E_1E_2D_R$ and $E_1D_RH_R$ models may both give non-significant chisquares. Some models may yield "nonsense" answers. We could immediately reject the $E_1D_RH_R$ model, for example, if we obtained a negative estimate of H_R. However, if all models give reasonable estimates, we would then reject any model with non-significant estimates. If two models give reasonable, significant estimates our only criterion for deciding which of the models is the correct one is to choose the model with the lowest chisquare. However, we must be very cautions in accepting the model with the lowest chisquare. This model will be the most likely explanation of variation in the trait, given the data available. However, it will not necessarily be the "correct" explanation of variation in the trait. In such a situation, further data should be collected in order to confirm the correct model of variation.

SECTION 4: ANALYSIS OF THE CORRELATIONS

1 INTRODUCTION

The correlation coefficient, a measure of the degree of association between two variables, was first formulated by Galton and perfected by Pearson, a mathematical statistician, nearly a hundred years ago and has been used in psychological studies ever since (e.g. Fisher, 1918; Hirsch, 1930; Newman, Freeman and Holzinger, 1937; Husen, 1959; Huntley, 1966; Burt, 1966). Its use antedates the analysis of variance and it is still used by psychologists despite its limitations. This is probably because of its simplicity especially in the analysis of the classical twin study and in all cases where experimental control cannot be exercised over the mating behaviour of the population under study (i.e. in human populations).

However, its limitations in the analysis of human behaviour far outweigh its advantages and Jinks and Fulker (1970) favour abandoning this approach altogether apart from in the reanalysis of previous studies. In the calculation of the intraclass correlation coefficient the between families variance, σ^2_b, is expressed as a fraction of the total variance and the total variance has thereby effectively been standardised to unity. Thus, in calculating the intraclass correlations of different groups of twins, we make the implicit assumption that their total variances are equal. If the total variances of the groups of interest are equal, there is no problem, but if they are not parameter estimates may be biased in a complex way, depending on the relative inequalities of the σ^2_T's present (Jinks and Fulker, 1970).

Effects leading to unequal variances will often be of genuine biological and psychological interest rather than mere anomalies. Thus, by standardising the total variances to unity, invaluable information may be lost. Jinks and Fulker (1970) list three failures of the simple model outlined in Section 3, which may lead to unequal total variances:

1. $G_1 + G_2 \neq G$

2. E_1's and E_2's not equal for all groups

3. Correlated environments.

Eaves (1976a, b) has recently formulated models for genotype-environment covariation of two types. Sibling effects (i.e. cooperation or competition between sibs) and cultural transmission (where the phenotypes of the parent forms a significant part of the environment of the offspring) both lead to distinct patterns of inequality in the total variances. Thus phenomena of interest may be obscured by use of the intraclass correlation coefficient. Parameters specifying gene actions, assortative mating and genotype-environment interactions may be included in a model if sufficient statistics are available for their estimation. However, effects leading to inequality of the total variances may not be included in a model for correlations and cannot be estimated. Thus the biometrical-genetical approach initiated by Fisher and developed by Mather and Jinks is superior to the correlational approach, since effects leading to unequal total variances may be specified in a model applicable to variance components and estimated.

However, we shall pursue the analysis of the correlation coefficients in order to compare the results of this study with those of previous studies.

2 HOMOGENEITY OF THE INTRACLASS CORRELATIONS

Intraclass correlation coefficients were calculated separately by race, sex and zygosity and are presented in Table A33. There are differences between the correlations of the eight groups and we must determine whether the correlations may be regarded as sample correlations from a population with a common value of r, with differences between them due to sampling variation alone, or whether the differences are significant and due to some underlying effect of interest.

A test for the homogeneity of correlations is given in all standard texts (e.g. Snedecor and Cochran). This test is sensitive to deviations from normality and since the distribution of r is skewed, we shall use a transformation, due to Fisher, of r into a quantity called z, which is distributed normally with a variance:

$$\sigma^2_z = 1/(N - 3/2)$$

in the case of the intraclass correlation coefficient. We may define z as follows:

$$z = \tfrac{1}{2} [\log_e (1 + r) - \log_e (1 - r)]$$

The correlations were transformed into z values and the test of homogeneity described by Snedecor and Cochran was carried out for each measure of ability. The results are presented in Table A34.

The correlations are heterogeneous for eight tests and further analysis is needed to elucidate the source of the heterogeneity. However,

TABLE A33: INTRACLASS CORRELATION COEFFICIENTS FOR THE EIGHT GROUPS OF TWINS ON ALL TESTS

	MONOZYGOTIC TWINS				DIZYGOTIC TWINS			
	Caucasian		Negro		Caucasian		Negro	
	Male	Female	Male	Female	Male	Female	Male	Female
Arithmetic	0.793	0.798	0.825	0.846	0.463	0.570	0.635	0.683
Vocabulary	0.540	0.507	0.435	0.401	0.339	0.174	0.216	0.221
Heim	0.876	0.822	0.796	0.756	0.471	0.649	0.478	0.528
Spelling	0.863	0.823	0.866	0.739	0.419	0.582	0.335	0.599
Calendar	0.511	0.440	0.542	0.530	0.352	0.417	0.369	0.430
Identical Pictures	0.753	0.765	0.408	0.516	0.535	0.561	-0.285	0.581
Cubes	0.551	0.264	0.193	0.315	0.362	0.221	-0.129	0.099
Surface Development	0.791	0.627	0.513	0.453	0.475	0.222	0.296	0.227
Formboard	0.655	0.465	0.412	0.027	0.557	0.351	0.460	0.235
Paper Folding	0.549	0.547	0.438	0.531	0.459	0.455	-0.217	0.124
Object Aperture	0.471	0.415	0.514	0.284	0.406	0.267	0.141	0.124
Spatial Ability	0.777	0.778	0.827	0.862	0.653	0.561	0.271	0.498
Mazes completed	0.414	0.623	0.727	0.403	0.527	0.293	0.240	0.162
Mazes errors	0.473	0.490	0.753	0.353	0.476	0.407	0.414	0.236

TABLE A34: RESULTS OF A TEST FOR THE HOMOGENEITY OF CORRELATIONS

TEST	χ^2_7	P	Significance level
Arithmetic	21.63	0.0029	**
Vocabulary	9.05	0.25	NS
Heim	29.17	0.0001	***
Spelling	34.09	0.00002	***
Calendar	1.91	0.96	NS
Identical Pictures	32.26	0.0004	***
Cubes	10.09	0.18	NS
Surface development	30.04	0.00009	***
Formboard	18.10	0.0115	*
Paper Folding	10.95	0.14	NS
Object Aperture	5.59	0.59	NS
Spatial ability	23.37	0.0015	**
Mazes completed	14.32	0.0458	*
Mazes errors	10.68	0.15	NS

it turns out that for six tests the correlations may be regarded as estimates of the same population value of the correlation between twins, differences between them being due to sampling error only. From the models discussed in Section 3, we recall that for a trait showing heritable variation, we expect the correlation between monozygotic twins to be greater than that for dizygotic twins. The simplest model for these six tests is that there are no differences between the correlations of monozygotic and dizygotic twins and, therefore, no evidence for a heritable component of variation in test performance. However, looking more closely at the data, we see small, but consistent differences between monozygotic and dizygotic twin correlations. Since the genetical model discussed in Section 3 predicts that r_{MZ} will be greater than r_{DZ}, we shall now proceed to make the best test of this expectation, which is not given by the test for overall homogeneity of correlations.

3 A MODEL FOR THE DIFFERENCES BETWEEN THE CORRELATIONS

In order to determine the source of the differences between the correlations, we shall adopt a similar approach to that used in Section 2.5 for the means. The simple linear model appropriate to the correlations was formulated in the same way as the model for the means. This model which recognises the effects of race, sex and zygosity and their interactions is presented in Table A35. We shall fit this model to the z values rather than the correlations since we wish to test the parameter estimates. We shall do this although transformation changes the scale because, although the method of weighted least squares described in Appendix E makes no assumptions about distribution during the estimation of the parameters and the variance-covariance matrix, normality has to be assumed in order to test the adequacy of the model and make tests of significance of its parameters.

The variance of z is given by:

$$\sigma_z^2 = \frac{1}{N - \frac{3}{2}}$$ where N is the number of pairs of twins. This means that the variance of each z differs since the number of pairs differs from group to group and, therefore, estimates of the effects derived from the linear model are not independent. Weighted least squares, using the inverses of the variances as weights, yields the maximum likelihood estimates of the parameters. These estimates and their standard errors, which are derived from the diagonal terms of the variance-covariance, matrix are presented in Table A36. An example of a variance-covariance matrix is presented in Table A37 for the

TABLE A35: SIMPLE LINEAR MODEL FOR TESTING DIFFERENCES IN THE CORRELATIONS

	MONOZYGOTIC				DIZYGOTIC			
	CAUCASIANS		NEGROES		CAUCASIANS		NEGROES	
	Males	Females	Males	Females	Males	Females	Males	Females
Mean	1	1	1	1	1	1	1	1
Zygosity	1	1	1	1	-1	-1	-1	-1
Race	1	1	-1	-1	1	1	-1	-1
Sex	1	-1	1	-1	1	-1	1	-1
ZxR	1	1	-1	-1	-1	-1	1	1
ZxS	1	-1	1	-1	-1	1	-1	1
RxS	1	-1	-1	1	1	-1	-1	1
ZxRxS	1	-1	-1	1	-1	1	1	-1

TABLE A36: EFFECTS CONTRIBUTING TO VARIATION IN THE Z VALUES OF TWINS FOR TESTS OF PERFORMANCE

Test	Main[1] effect	Zygosities[1] effect	Race effect	Sex effect	Zygosity x race	Zygosity x sex	Race x sex	Zygosity x race x sex
Arithmetic	0.915***	0.232***	-0.085NS	-0.039NS	0.024NS	0.019NS	-0.001NS	0.015NS
Vocabulary	0.378***	0.135*	0.045NS	0.032NS	0.023NS	-0.011NS	0.023NS	-0.022NS
Heim	0.874***	0.275***	0.078NS	-0.044NS	0.034NS	0.078NS	-0.013NS	0.036NS
Spelling	0.839***	0.301***	0.057NS	-0.029NS	0.039NS	0.111NS	0.009NS	-0.022NS
Calendar	0.487***	0.072NS	-0.025NS	-0.005NS	-0.016NS	0.032NS	0.009NS	0.010NS
Identical Pictures	0.574***	0.174**	0.231***	-0.145*	0.015NS	0.104 +	0.129*	-0.101NS
Cubes	0.248***	0.105NS	0.125 +	0.018NS	-0.033NS	0.037NS	0.108 +	0.012NS
Surface Development	0.518***	0.198**	0.120 +	0.098NS	0.069NS	0.006NS	0.059NS	0.005NS
Formboard	0.436***	0.003NS	0.135*	0.151**	0.071NS	0.021NS	-0.016NS	-0.017NS
Paper folding	0.398***	0.175**	0.157*	-0.057NS	-0.114 +	0.028NS	0.059NS	-0.028NS
Object aperture	0.348***	0.105 +	0.066NS	0.065NS	-0.043NS	0.021NS	-0.008NS	-0.043NS
Spatial ability	0.850***	0.290***	0.036NS	-0.031NS	-0.124 +	0.000NS	0.067NS	-0.037NS
Mazes completed	0.477***	0.153*	0.038NS	0.071NS	-0.082NS	-0.020NS	-0.073NS	-0.123 +
Mazes errors	0.276***	0.155*	0.004NS	0.012NS	0.003NS	0.085NS	-0.032NS	-0.015NS

The standard error of every estimate is 0.065

Key

NS Not significant at the 10% level
\+ Significant at the 10% level
* Significant at the 5% level
** Significant at the 1% level
*** Significant at the 0.1% level

[1] the significance levels for the one-tailed test are given for these effects, since we have an a priori expectation that these will be positive (see text)

TABLE A37: THE VARIANCE-COVARIANCE MATRIX FOR THE ARITHMETIC TEST

	Mean	Zygosity	Race	Sex	ZxR	ZxS	RxS	ZxRxS
Mean	42	-11	-20	16	7	-7	-12	4
Zygosity	-11	42	7	-7	-20	16	4	-12
Race	-20	7	42	-12	-11	4	16	-7
Sex	16	-7	-12	42	4	-11	-20	7
ZxR	7	-20	-11	4	42	-12	-7	16
ZxS	-7	16	4	-11	-12	42	7	-20
RxS	-12	4	16	-20	-7	7	42	-11
ZxRxS	4	-12	-7	7	16	-20	-11	42

Note: All figures in the body of the Table must be multiplied by 10^{-4} to obtain the true values

Arithmetic Test. This matrix is derived from the weight and model matrices and is the same for every test since only the vector of statistics varies from test to test. We see that the diagonal terms are identical because every z enters into every comparison. Thus, each estimate has the same standard error.

Inclusion of the mean as an effect provides a test of the homogeneity of the z values. However, we are proceeding with no test for equality of the total variances, which could lead to unsuspected biases in estimates of genetical and environmental effects. Since we are interested only in the significance of the effects and not their absolute size in this analysis, we shall not be seriously misleading outselves by our failure to test this assumption.

4 THE EFFECT OF ZYGOSITY

4.1 Tests showing Evidence of Heritable Variation

The simple genetical model for correlations predicts that the correlation between pairs of monozygotic twins will be greater than the correlation between dizygotic twins. Therefore, if there is a heritable component of variation for the abilities measured here, we expect the zygosities effect to be significant, and positive. Since we expect the effect to be positive, a one-tailed test of significance is appropriate for the zygosities effect. The significance levels obtained using the one-tailed test are shown in Table A36.

The zygosities effect is significant at the five per cent level at least, for ten of the fourteen tests and approaches significance for Cubes and Object Aperture. We notice a significant zygosities effect for several tests whose correlations were homogenous using the test for overall homogeneity of correlations described in the last section. We conclude that differences between monozygotic and dizygotic twin correlations must be small for these tests, since the test for overall homogeneity of correlations did not detect them. However, they must be consistent since a significant zygosities effect was found in this analysis.

Thus, there is evidence for a heritable component of variation for twelve of our fourteen tests. Of the two remaining tests, for which we have no evidence for a genetical component, one, the Formboard Test, belongs to the Spatial Factor. The other is the Calendar Test which loads about equally on the Verbal-Arithmetic Factor and the Spatial

Factor. It is surprising that these tests show no evidence of heritable variation since they measure similar abilities and load on the same Factor as the other Spatial tests which do show some evidence for heritable variation. We notice also that the zygosities effect for two other tests belonging to the Spatial Factor i.e. Cubes and Object Aperture, only approach significance. It seems more difficult to detect heritable variation for Spatial than for Verbal-Arithmetic Tests. Looking at Table A33, the table of intraclass correlations, we see that, overall, the correlations are higher for tests of the Verbal-Arithmetic Factor than for those of the Spatial Factor. We also see a correspondence between significance level and size of correlation. Those tests with monozygotic twin correlations over 0.70 all show a highly significant zygosities effect, whereas those with correlations less than this either have zygosities effects significant at the five per cent level only or no significant zygosities effect at all.

The lower monozygotic twin correlations for tests of the Spatial Factor imply that E_1 is larger for the Spatial than for the Verbal-Arithmetic Tests. This might make any component of between families variation more difficult to detect for the Spatial Tests. Thus, our problem is not that some Spatial Tests have a heritable component of variation and others do not, but that some have a higher E_1 than others. The heritability is thus reduced as is the power of the test for detecting genetical variation.

4.2 Power of the Test for Detecting Heritable Variation

We may obtain the probability of correctly rejecting the Null Hypothesis of no difference between monozygotic and dizygotic twin correlations by a few simple power calculations.

The lower this probability, the more likely we are to retain the Null Hypothesis even though it is false. We may also calculate the number of pairs of twins needed, to detect a significant zygosities effect, in ninety five per cent of cases, given the standard errors and estimates from this analysis.

Probability of correctly rejecting the Null Hypothesis

If there is a heritable component of variation in ability we expect z for monozygotic twins to be greater than z for dizygotic twins. Therefore, the test of the difference between the transformed correlationsis:

$$C = (z_{MZ} - z_{DZ}) / \sigma(z_{MZ} - z_{DZ})$$

and we expect C to be greater than or equal to 1.65. The value of $z_{MZ}-z_{DZ}$ depends on the heritability and the genetical system governing the trait (Eaves and Jinks,1972). The value of the standard deviation of the difference depends only on sample size, since we are considering z values. Therefore, given certain assumptions about the sizes of these variables, we can calculate the power of the test by integrating the area under the normal curve, N(0,1), between the limits (1.65 - C) and infinity.

Given that there are no twins reared apart in this study, an estimate of the broad heritability is given by r_{MZ}. This estimate is of course inflated by E_2. Therefore, assuming no E_2, we took values of heritability from 0.40 to 0.90 at 0.10 intervals, since in our study the monozygotic twin correlations varied between these two limits. We now considered two cases in each instance assuming no dominance. In the case of a random mating population the expected correlation of dizygotic twins is half that of monozygotic twins. Assuming a correlation of 0.5 between the genetical deviations of spouses, then the expected value of the dizygotic correlation is three quarters that of the monozygotic correlation. We calculated the expected dizygotic correlations, corresponding to the six monozygotic correlations we are using, both for the case of random mating and for that of assortative mating. We then transformed both monozygotic and dizygotic twin correlations into z values and found the expected differences between the monozygotic and dizygotic twin correlations. To find the expected value of C, we must now calculate the appropriate standard deviations. The standard deviation is the square root of the pooled variance of monozygotic and dizygotic twins:

$$S.D. = \sqrt{\frac{1}{N_{MZ} - \frac{3}{2}} + \frac{1}{N_{DZ} - \frac{3}{2}}}$$

where N_{MZ} and N_{DZ} are the number of pairs of twins on which the monozygotic and dizygotic twin correlations were based. We shall use the numbers from this study and thus find the power of the test for the four race-sex groups. The expected values of C are the expected differences between z values, based on the six heritability values

divided in turn by the standard deviation of each of the four groups. The power of the test is given by the area under the normal curve between (1.65-C) and infinity, which was found from standard tables. The powers of the tests which we have calculated are presented in Tables A38, parts A and B. We shall adopt as our criterion of a "good" test that we should correctly reject the null hypothesis of no difference between the correlations in 95 per cent or more cases. On this criterion, we have a "good" test with the sample sizes used in this study only if the monozygotic twin correlation is 0.8 or higher for Whites assuming random mating, or 0.9 or more in all other cases. With the correlations of around 0.50 found for several Spatial tests, the power of the test is below 54 per cent in all cases and is less than 20 per cent if there is assortative mating. Since there is evidence of assortative mating for ability from other sources (e.g. Vandenberg, 1967) the power of our test for detecting heritable variation is very low for most of the spatial tests. Therefore, we shall pursue the genetical analysis of all these tests of ability, recognising that we are unlikely to detect heritable variation for several of the Spatial Tests.

The number of pairs needed to detect a heritable component of variation in 95 per cent of cases

The number of pairs required to detect heritable variation in 95 per cent of cases, for the situations described in the last section is given by:

$$N = \frac{(C_e)^2}{(C_o)^2} \times n$$

TABLE A38: POWER OF THE TEST FOR A HERITABLE COMPONENT OF VARIATION IN ABILITY

PART A: A RANDOM MATING POPULATION

			HERITABILITY					
		n	0.4	0.5	0.6	0.7	0.8	0.9
White	Males	108	0.291	0.429	0.603	0.802	0.957	0.999
	Females	143	0.367	0.536	0.732	0.905	0.990	$\simeq$1.000
Black	Males	34	0.142	0.192	0.264	0.371	0.552	0.829
	Females	79	0.239	0.348	0.496	0.688	0.891	0.995

PART B: A POPULATION WITH A COEFFICIENT OF ASSORTATIVE MATING OF 0.50

			HERITABILITY					
		n	0.4	0.5	0.6	0.7	0.8	0.9
White	Males	108	0.140	0.192	0.271	0.409	0.644	0.945
	Females	143	0.164	0.233	0.337	0.512	0.733	0.986
Black	Males	34	0.090	0.107	0.136	0.184	0.302	0.528
	Females	79	0.123	0.161	0.224	0.330	0.532	0.873

where N is the number of pairs needed

n is the number of pairs in this study

C_e is the value of c needed for the one-tailed test to be significant at the 5 per cent level in 95 per cent of cases.

C_o is the value of c for the hypothetical differences between r_{MZ} and r_{DZ} for the cases described in the previous section.

The number of pairs required to detect a heritable component of variation for the cases used in the previous section were calculated and are presented in Table A39. This number is always greater than the number of pairs in this study unless the heritability is 0.80 or greater. Therefore, the chance of detecting a heritable component of variation for those tests having a monozygotic twin correlation less than 0.80 is low and it is not surprising that we fail to detect a significant zygosities effect for several of our tests (cf. Eaves and Jinks (1972)).

4.3 Possible Causes of Differences Observed between Tests

We have seen that the tests fall into two groups corresponding to the Factors identified in Section 2.1, according to the size of the monozygotic twin correlations and the presence of a significant zygosities effect. Tests of the Verbal Factor have higher correlations and show significant zygosities effects, whereas the tests of the Spatial Factor tend to have lower correlations and the zygosities effect is not always significant. Differences between the monozygotic twin correlations implies some variation in the proportion of variation

TABLE A39: NUMBER OF PAIRS OF TWINS NEEDED TO DETECT A HERITABLE COMPONENT OF VARIATION

PART A: A RANDOM MATING POPULATION

			HERITABILITY					
		n	0.4	0.5	0.6	0.7	0.8	0.9
White	Males	108	968	543	321	188	104	48
	Females	143	913	512	303	177	98	46
Black	Males	34	1092	612	362	212	117	55
	Females	79	965	541	320	187	103	48

PART B: A POPULATION WITH A COEFFICIENT OF ASSORTATIVE MATING OF 0.50

			HERITABILITY					
		n	0.4	0.5	0.6	0.7	0.8	0.9
White	Males	108	3629	1964	1089	586	288	111
	Females	143	3430	2896	1027	553	271	105
Black	Males	34	4094	2215	1228	661	324	125
	Females	79	3617	1957	1085	584	287	111

attributable to E_1 between tests. Tests of the Verbal-Arithmetic Factor have a smaller E_1. The Spatial Tests show a range of correlations and, therefore, a range of E_1's, varying from an E_1 value almost as small as that for the Verbal tests, to one which accounts for most of the variation. Within families environmental variation can be produced by a number of factors. Accidents during development and experiences specific to one member of a pair produce E_1 differences. It seems unlikely that these should have a differential effect on tests of spatial ability. However, test unreliability also produces variance within pairs. It seems likely that the larger E_1 of some spatial tests may be caused by the greater unreliability of these tests.

It is interesting that the Verbal Tests turn out to have higher monozygotic twin correlations. This duplicates findings of other workers (e.g. Conrad and Jones, 1940). Verbal tests are generally assumed to be culture-biased, requiring some knowledge of language and social customs. Therefore, differences between individuals in performance on these tests might be expected to have a large environmental component due to environmental differences between families. Conversely Spatial tests seem to require no prior "knowledge" and are assumed to be more "culture fair". Therefore, environmental factors should be less important. Since monozygotic twin correlations include variance due to genetical influences and also between families environmental influences (E_2) it is possible that the correlations for Verbal Tests may be higher because they are boosted by E_2 effects. We will make some attempt to set an upper limit for E_2 in later sections, but we cannot estimate it directly without data on relatives reared apart.

5 THE RACES EFFECT AND THE ZYGOSITY X RACES INTERACTION

We have no expectations about whether a races effect or a zygosities x races interaction will be positive or negative. Therefore, a two-tailed test of significance was made for these effects, and is shown in Table A36. The races effect is small and non-significant for the Verbal-Arithmetic tests. However, there is a significant positive races effect for Identical pictures and Formboard and this effect approaches significance for Cubes and Surface development. This reflects the overall tendency, which can be seen from Table A33, for the correlations of Whites to be slightly higher than the correlations for Blacks. This indicates either that E_1 may be smaller in Whites than Blacks for those tests with a significant races effect, or alternatively, between families differences (B) either genetical or environmental, may be more important in Whites. This would accord with the observation of greater heterogeneity in Whites between sub-groups of the data, recorded in Section 1.5.3. However, we can only offer these suggestions, recognising that if there is heterogeneity between the total variances of Whites and Blacks then it is not valid to make comparisons of variance components in Whites and Blacks from the intraclass correlations.

There have been reports that the heritability of intelligence is higher in Whites than in Blacks (e.g. Scarr-Salapatek, 1971). These reports have generated some controversy and we shall examine the test for a race difference in heritability inherent in this analysis. Such a difference implies that the magnitude of any heritable component depends on race i.e. that there is a zygosities x races interaction. We see from Table A36 that there is no evidence for such an interaction

in these data. If there were an interaction, which was non-significant because the power of the test was too low, then we would expect the interaction to take the same sign for all tests. However, the interaction takes both signs with almost equal frequency and thus there is not even any evidence for a directional trend in these data. We conclude that there is no evidence for a difference in heritability between the races. This confirms the results of the analysis of Eaves and Jinks (1972) using data on White and Black twins from Philadelphia.

6 THE SEX EFFECT AND THE ZYGOSITY X SEX INTERACTION

6.1 Specification of Sex Linkage

Several workers have argued that the mean score of males on tests of spatial ability is higher than that of females because a substantial proportion of variation in Spatial Ability is produced by a single sex-linked recessive gene (Stafford, 1961; Hartlage, 1970; Bock and Kolakowski, 1973).

There is a test of this hypothesis in these data. In the absence of sex-linkage, we expect the variance of males and females to be the same. The total variance for a random mating population is defined by Mather and Jinks (1971) as follows:

$$\sigma^2_T = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R + E_1 + E_2$$

Mather and Jinks (1963) show that, with sex-linkage, the expected variances of males and females respectively, given random mating, are:

$$\sigma^2_{T_m} = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R + E_1 + E_2 + D'_x$$

$$\sigma^2_{T_f} = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R + E_1 + E_2 + \tfrac{1}{2}D_{R_x} + \tfrac{1}{4}H_{R_x}$$

where D'_x is the additive variance due to the sex-linked genes in males

$\tfrac{1}{2}D_{R_x}$ is the additive variance due to the sex-linked genes in females

$\tfrac{1}{4}H_{R_x}$ is the dominance variation due to the sex-linked genes (in females)

Therefore, sex-linkage may lead to inequality of total variances between the sexes in certain circumstances. If there is no dominance for the trait and no sex limitation (i.e. if $D'_x = D_{R_x}$), then the variance of males will be higher than that of females. The presence of dominance will reduce the difference between the total variances of the two sexes. Sex limitation will change the relative contribution of the sex-linked genes to the total variances and may reduce the difference in total variance between the sexes. Thus, a difference in variance between males and females is diagnostic of sex-linkage but homogeneity of the variances does not necessarily mean there is no sex-linkage. In this Section we are concerned only with the correlations, and the properties of the total variances will be discussed further in the next section.

The correlations for monozygotic and dizygotic twins are given by:

$$\text{Correlation MZ males} = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R + E_2 + D'_x/\sigma^2_{T_m}$$

$$\text{Correlation MZ females} = \tfrac{1}{2}D_R + \tfrac{1}{4}H_R + E_2 + \tfrac{1}{2}D_{R_x} + \tfrac{1}{4}H_R/\sigma^2_{T_f}$$

$$\text{Correlation DZ males} = \tfrac{1}{4}D_R + \tfrac{1}{16}H_R + E_2 + \tfrac{1}{2}D'_x/\sigma^2_{T_m}$$

$$\text{Correlation DZ females} = \tfrac{1}{4}D_R + \tfrac{1}{16}H_R + E_2 + \tfrac{3}{8}D_{R_x} + \tfrac{1}{16}H_{R_x}/\sigma^2_{T_f}$$

Now, providing the environmental components and the rest of the genetical system are the same in the two sexes, we can make some predictions about the effects expected in this analysis of the correlations, given sex-linkage. Firstly, assuming no dominance or sex-limitation, we expect the DZ twin correlation to be higher in females than males since $\tfrac{3}{8}D_{R_x}/\tfrac{1}{2}D_{R_x} > \tfrac{1}{2}D'_x/D'_x$. We expect the MZ

twin correlations in males to be higher than those in females since $D'_x/D'_x + E_1 > \frac{1}{2}D_{R_x}/\frac{1}{2}D_{R_x} + E_1$ if the gene effects are the same in males and females.

Thus, we will only find a significant sex effect if the effects of sex-linkage on the MZ and DZ correlations do not cancel out. This will of course depend on the environmental components. So, although a sex effect may indicate sex-linkage, its absence does not necessarily imply no sex-linkage. Comparing the differences between the correlations of monozygotic and dizygotic twins in males and females, we see that we expect the difference to be $\frac{1}{2}D_{R_x}$ in males and only $\frac{1}{8}D_{R_x}$ in females. Thus the magnitude of the zygosities effect will depend on sex and we expect, in theory, a positive zygosities x sex interaction if there is sex-linkage. Dominance and sex limitation will not affect the relative values of the correlation coefficients much, since they contribute to σ^2_b and σ^2_T and, therefore, we expect to find a sex effect and zygosities x sex effect even in the presence of dominance and sex limitation, if there is sex-linkage.

The sensitivity of this test for sex linkage depends on the proportion of the total variation accounted for by the sex-linked genes. The sensitivity of the test will be highest when all variation in Spatial Ability is produced by sex-linked genes. However, given that this is not the case, then the sensitivity will depend on the relative sizes of $\frac{1}{2}D_R$ and $(D'_x, \frac{1}{2}D_{R_x})$ and also if there is dominance of $\frac{1}{4}H_R$ and $\frac{1}{4}H_{R_x}$. The larger $\frac{1}{2}D_R$ compared with $(D'_x, \frac{1}{2}D_{R_x})$, the lower the sensitivity of the test will be and, of course, the less important will be the contribution of sex-linked genes to variation in Spatial Ability.

A difference between male and female correlations may indicate sex-linkage. Such a difference, however, could be produced by other causes such as different E_1's in the two sexes i.e. σ^2_T could be greater for one sex than the other. Therefore, a sex effect although it might imply sex-linkage, cannot be taken as diagnostic of sex-linkage.

The zygosities x sex interaction does not depend on the environmental effects being the same in the two sexes and is, therefore, diagnostic of sex-linkage.

6.2 Evidence for Sex-Linkage in these Data

Looking at Table A36, we see that the sex effect is positive in half the cases and negative for the other half, and although the sex effect is significant at the five per cent level for two tests, it is positive in one case and negative in the other. This provides no evidence for a sex effect in either direction in these data.

Different environmental components of variation in the two sexes may obscure any sex effect due to sex linkage, although it could also produce a sex effect in the absence of sex-linkage. The two significant sex effects for Identical Pictures and Formboard might be produced by different E_1's in the two sexes. It seems likely that this would be caused by differential test reliability in the two sexes, since there was no evidence for such an effect for any other Spatial Test. We also suggested that these tests may have a low reliability in Section 4.5.3.

A positive significant zygosities x sexes interaction is our best diagnostic of sex-linkage. We see from Table A36 that this interaction is not significant, that the size of the effect is small in all cases and is negative for several tests. Therefore, we have no evidence at all suggestive of sex-linkage.

Of course, our failure to detect a zygosities x sexes interaction is not proof of no sex-linkage, but only of our failure to detect it. The power of our test could be too low to pick up the effects of sex linkage, especially if the contribution to variation in Spatial Ability of sex-linked genes is small. The Null Hypothesis in testing the zygosities x sexes interaction is that $(z_{MZ} - z_{DZ})$ males - $(z_{MZ} - z_{DZ})$ females, is zero. The variance of this difference is: $\sigma^2 = \sigma^2 (z_{MZ} - z_{DZ})$ males + $\sigma^2 (z_{MZ} - z_{DZ})$ females which is 0.049059 for the sample sizes used in this study. We expect this difference to be positive and, therefore, we shall reject the null hypothesis at the five per cent level if this difference divided by the standard error is greater than 1.65.

For the power calculations we shall assume that all the variation in Spatial Ability is due to sex-linked genes, that there is no dominance and that $D'_x = D_{R_x}$ (i.e. no sex limitation). We shall consider two cases. Firstly, we shall take the following values for our parameters:

$D_R = 0.60$

$E_1 = 0.20$

$E_2 = 0.20$

This gives us correlations as follows:

Correlation of MZ males = 0.80

Correlation of MZ females = 0.71

Correlation of DZ males = 0.50

Correlation of DZ females = 0.64

After transforming to z, it turns out that $(z_{MZ} - z_{DZ})_{males} - (z_{MZ} - z_{DZ})_{females}$ equals 0.4202. Therefore, c = 0.47844/0.22149 = 1.90. Now, taking (1.65 - 2.16) and finding the area under the normal curve of unit variance between the limits of -0.25 and infinity, the power of the test is 0.60. Now taking new values of our parameters of $D_R = 0.40$, $E_1 = 0.30$, $E_2 = 0.30$, we can calculate the power in exactly the same way and it turns out to be 0.26 in this case. Thus even in two extreme cases where all the variation in Spatial Ability is produced by sex-linked genes and a large proportion of the variance is due to additive effects, we are unlikely to detect a significant zygosities x sexes interaction with samples of this size. To be certain of detecting a significant zygosities x sexes interaction in 95 per cent of cases, 1098 pairs of twins would be needed for the first example and 3972 pairs for the second example, given the sample structure used here. Since we have used extreme examples, it seems unlikely that a study of this design (i.e. monozygotic and dizygotic twins reared together) could demonstrate sex-linkage unless prohibitively large sample sizes were used.

6.3 Comparison of the Test for Sex-Linkage in these Data with other Tests for Sex-Linkage

More usual tests for sex-linkage are provided by the use of full-sib and parent-offspring correlations. These might provide more powerful tests for sex-linkage than the use of monozygotic and dizygotic twins reported in this Study.

The expectations for full-sib correlations are, of course, the same as those for dizygotic twins reared together and, therefore, we expect the correlations of females to be higher than those of males. We can, of course, test whether the correlation of females is higher than that of males in these data. However, since the correlations of males are higher than those for females for fourteen of the eighteen comparisons we can make using the correlations for the Spatial Tests, there was no point in trying this test with these data. This is further evidence for the absence of sex-linkage in these data, although it cannot be taken as conclusive since the total variances and the environmental components of variation may not be the same in the two sexes. Also we have assumed no dominance or sex limitation in formulating this test.

We can calculate the power of this test using the sample sizes of this Study. Mather and Jinks (1963) show that the maximum value the correlation can take in males is 0.50 whereas it is 0.75 in females, assuming no dominance or environmental variation. After transforming these correlations into z values, the difference between them was 0.4236 giving a value of c of 2.50. Thus, the power of the test, calculated in the usual way is 0.80. This is an overestimate of the power of the test in these data, since no dizygotic twin correlation was

as high as 0.75 and we assumed all variation to be due to sex-linked genes. The normal test of sex-linkage available in these data is thus not remarkably more powerful than the test we have done.

Mather and Jinks (1971) give the expectations for parent/offspring covariances in the presence of sex linkage and show that given no sex-limitation, the relative magnitudes of the possible covariances will be: female/male = male/female $\geqslant$ female/female $>$ male/male. Since in this case we are using covariances, we have no theoretical expectations for the standard deviations of these comparisons and we cannot do power calculations unless parent/offspring data is available. However, if we had data on all four types of parent/offspring covariance, then the ordering of the correlations as above will be highly indicative of sex-linkage even if significant differences could not be found using sample sizes comparable to those used in this study. Thus, parent/offspring data would be more useful for the detection of sex-linkage than data on monozygotic and dizygotic twins.

We conclude that although we do not have the best data set for the detection of sex-linkage, there is no evidence whatsoever to suggest its presence in these data. Indeed, if we assume that environmental components of variation for males and females are the same, then the fact that the male correlations for dizygotic twins are higher than those of females in nearly all cases is evidence for no sex-linkage. Thus, in order to explain the higher mean of males in tests of Spatial Ability, we must invoke either that the environmental components of variation differ between the sexes or that there is sex-limitation of Spatial Ability.

7 SUMMARY

Overall there is little evidence for any consistent, significant effect of race or sex or interactions involving them on the correlations. This does not support previous suggestions of:

1. A difference in heritability between Whites and Blacks
2. Sex-Linkage

although the power of our tests was low.

There was a marked difference between MZ and DZ correlations, which was significant for most tests. This confirms previous work suggesting heritable variation for measures of ability. The heritable variation expressed as a proportion of the total was less marked for the Spatial than the Verbal-Arithmetic tests, and consequently the power of the test for detecting the heritable component was lower. Thus the zygosities effect was not significant in several cases.

We suggested that greater unreliability of the Spatial tests might be inflating E_1 and thus reducing the relative importance of the heritable component for these tests.

Very little useful information can be gained from further analysis of the correlations and we turn now to an analysis of the variance components in order to make the most efficient use of the information available in the data and to explore further the questions raised by the analysis described in this Section.

SECTION 5: ANALYSIS OF THE VARIANCE COMPONENTS

1 INTRODUCTION

The limitations of the correlational approach have been discussed. In calculating correlation coefficients, total variances are standardised to unit variance. The assumption that the total variances are homogeneous is made implicitly in calculating the correlations. Usually no test of this assumption is made before estimating parameters from the correlations, so if the total variances were heterogeneous, parameter estimates would be biased in a complex way and rendered meaningless. Another limitation is that only effects not leading to inequality of total variances may be specified in models for correlations. Parameters specifying genotype-environmental covariation, for example, cannot be estimated efficiently from correlations. Several tests for homogeneity of total variances are available (Winer, 1962) but Bartlett's test is the most sensitive and is commonly used. If the total variances turn out to be homogeneous, unbiased estimates of genetical and environmental components of variation may be obtained from the correlations of the appropriate groups. Failure of the assumption of homogeneity gives no indication of the source of the heterogeneity. However, it shows that analysis of the correlations is worthless and that the source of the heterogeneity must be sought by an analysis of variance components.

When models are fitted to correlations, heterogeneity of total variances has been removed by standardisation and inappropriate models may fit the data. However, when models are fitted to variances,

failure of the assumption of homogeneity leads to failure of the model and thus heterogeneity will be detected. When heterogeneity is found, parameters specifying effects leading to inequality of total variances can be specified for the variance components and such models can be tested for adequacy. Thus, model fitting to variance components is far more useful than model fitting to correlations, since it provides a test of the assumption of homogeneity of total variances and allows us to fit more complex models when appropriate.

In this Section the outcome of fitting models to the between and within pairs mean squares of the fourteen measures of ability will be described. Eventually a fairly simple picture, consistent over tests, will emerge. However, the results are complex in their detail and several anomalies occur. Therefore, a brief outline of the findings is given here so that the peculiarities of individual tests may be taken in context, as they occur.

Significant differences between MZ and DZ correlations were found for Verbal-Arithmetic tests and simple environmental models were not adequate to explain the observed pattern of variation. The simple additive genetical model did fit the data.

The effect of zygosity on the correlations was not significant for several Spatial tests. Here both simple environmental and simple additive genetical models were adequate to explain variation. Where the zygosities effect was significant the simple additive genetical model was required to explain the data.

Difficulty in discriminating between alternative models for several spatial tests could be explained in terms of the power of the test. The biased broad heritability of these tests was lower, possibly

because unreliability accounted for a greater proportion of the total variation. This would make it harder to detect the genetical component. Overall, the data suggested a heritable component for both Verbal-Arithmetic and Spatial tests, although some problems were found.

Heterogeneity between races was suggested, particularly for Spatial tests, by a test for homogeneity of total variances and models had to be fitted separately to Whites and Blacks in some cases. Heterogeneity between sexes was found for several Spatial tests. Estimates of biased broad heritability were compared between races and sexes, but did not differ significantly. Generally,Whites were more variable than Blacks and males more variable than females. The difference in variability between groups could not be ascribed to any particular component of the model during these analyses since all components differed between groups.

Comparison of estimates of $\hat{B}$ and $\hat{G}_1$ provided evidence for the role of either common environmental effects or assortative mating and an attempt was made to set upper and lower limits to these components, using estimates of the marital correlation obtained from other studies. Assuming no dominance or assortative mating estimates of $\hat{E}_2$ were obtained. Narrow heritabilities were subsequently calculated and found to be non-significantly greater in Whites than Blacks. The estimates were substantially higher for Verbal-Arithmetic tests than for Spatial tests. It was tentatively concluded that common environmental effects (or additive genetical variation produced by assortative mating) account for a greater proportion of the total variation in Blacks. The data indicated that these effects were more important in determining Spatial than Verbal-Arithmetic ability.

Discussion of these findings in relation to the observed heterogeneity between the three sub-groups of the data was postponed so that the results for specific abilities and general ability could be discussed simultaneously.

2 TESTING THE ASSUMPTION OF HOMOGENEITY OF TOTAL VARIANCES

2.1 Testing for Homogeneity

The mean squares for the data of the Georgia Twin Study were presented in Table A24. Bartlett's test (Bartlett, 1937) was used to test the null hypothesis that these eight mean squares are estimates of the same σ^2. The chisquares testing this hypothesis for each measure of ability are shown in Table A40. Spelling shows heterogeneity of total variances, significant at the five per cent level and chisquare approaches significance for Vocabulary. Of the Spatial Tests, five show highly significant heterogeneity. Chisquare for the Calendar Test approaches significance. For these tests showing some evidence of heterogeneity of total variances, it is not appropriate to fit models to all groups jointly.

Heterogeneity of total variances may be produced by differences in variance between races, sexes, or zygosities. Each of these effects, if present, would alter the model fitting procedure. Therefore, the source of the heterogeneity must be determined so that we can decide which models to fit and to which groups to fit them.

We shall adopt the approach employed in Sections 2.5 and 4.4 to formulate a linear model specifying the effects on the variances of race, sex, zygosity and their interactions. The model is identical to that shown in Table A25 for the means. Before fitting this model, the variances were transformed by taking their logarithms. The log variances have a theoretical error which is independent of the variance and depends only on the sample size. It can be shown (Gale, personal comm.)

TABLE A40: RESULTS OF BARTLETT'S TEST FOR THE HOMOGENEITY OF TOTAL VARIANCES

Test	χ^2_7	P	Significance level
Arithmetic	6.54	30-50	NS
Vocabulary	13.07	5-10	†
Heim	6.10	50-70	NS
Spelling	15.91	2-5	*
Calendar	13.88	5-10	†
Identical pictures	6.72	30-50	NS
Cubes	29.08	< 0.001	***
Surface development	80.39	< 0.001	***
Formboard	77.99	< 0.001	***
Paper folding	11.23	10-20	NS
Object aperture	76.76	< 0.001	***
Spatial ability	28.02	< 0.001	***
Mazes completed	9.09	20-30	NS
Mazes errors	6.66	40-50	NS

KEY:

NS	Not significant
†	Significant at the 10% level
*	Significant at the 5% level
**	Significant at the 1% level
***	Significant at the 0.1% level

that the variance of the log variance is given by

$$V_{(\log \text{ variance})} = \frac{2}{N-1} \quad \text{where N is the sample size}$$

The model was fitted to the log variances by weighted least squares, using the inverses of $V_{(\log \text{ variance})}$ as the weights. The estimates, their standard errors and significance levels are given in Table A41.

2.2 Inequality of Variances between Zygosities

If the total variances of monozygotic and dizygotic twins are heterogeneous, one or more of the assumptions on which the simple genetical model is based have failed and it is not appropriate to use this model. Monozygotic and dizygotic twins belong to the same population and we expect genetical variation to be the same in the two types of twin. However, several factors, described in Section 3.2, may lead to inequality of the variances of monozygotic and dizygotic twins. Environmental components of variation may not be the same in the two groups. The environment specific to individuals (E_1) might differ because, for example, of differences in *in utero* environment, or in treatment of the two types of twin. Both between and within families environmental components of variation could differ between monozygotic and dizygotic twins if there were genotype-environment interaction. However, from Table A41, we see that there is no evidence for heterogeneity of the total variances of the two types of twin. If there had been heterogeneity we would need to collect data on additional types of family, so that new models could be specified including parameters for

TABLE A41: EFFECTS CONTRIBUTING TO HETEROGENEITY OF TOTAL VARIANCES BETWEEN GROUPS OF TWINS

Test	Mean	Race	Sex	Zygosity	RxS	RxZ	SxZ	RxSxZ
Arithmetic	6.493***	-0.063NS	-0.029NS	0.031NS	0.071NS	-0.065NS	0.088NS	-0.106NS
Vocabulary	2.205***	0.212**	-0.104NS	0.058NS	0.101NS	0.018NS	0.006NS	-0.050NS
Heim	5.710***	0.011NS	0.123NS	0.031NS	0.030NS	-0.038NS	-0.039NS	0.025NS
Spelling	5.229***	-0.149*	0.060NS	0.049NS	0.102NS	-0.055NS	0.038NS	0.021NS
Calendar	3.596***	0.086NS	0.008NS	0.134NS	0.117NS	-0.064NS	0.050NS	0.057NS
Identical pictures	5.487***	-0.101NS	-0.050NS	0.027NS	0.101NS	0.040NS	-0.037NS	-0.015NS
Cubes	4.073***	0.250***	0.132NS	0.120NS	-0.030NS	-0.154*	-0.085NS	0.081NS
Surface development	4.508***	0.429***	0.242***	-0.021NS	0.014NS	0.042NS	0.036NS	-0.020NS
Formboard	3.274***	0.483***	0.126NS	0.058NS	-0.045NS	0.035NS	0.055NS	0.052NS
Paper folding	2.765***	0.125NS	0.042NS	0.123NS	0.132NS	-0.078NS	0.005NS	-0.035NS
Object aperture	3.230***	0.329***	0.343***	-0.052NS	-0.012NS	0.059**	-0.164NS	0.104NS
Spatial ability	5.558***	0.214**	0.160*	0.118NS	-0.067NS	0.203NS	-0.000NS	-0.035NS
Mazes completed	4.358***	0.020NS	-0.040NS	-0.054NS	0.052NS	-0.067***	-0.164NS	-0.012***
Mazes errors	4.418***	-0.286***	-0.266***	0.234***	0.388***	-0.273***	0.125NS	-0.285***

NOTE: The standard error, which is the same for all estimates, is 0.070

different E_1's and E_2's for MZ and DZ twins and perhaps for the effects of genotype-environment interaction. Without collection of new data the biases introduced by fitting the simple model would have to be recognised and taken into account during interpretation of the model.

Genotype-environment covariation could also produce inequality of the total variances. Covariation caused by competition or cooperation between sibs leads to distinctive patterns of inequality in the total variances, which, assuming equality of the environmental components, we could detect. (Eaves 1976a, b). Again we could either collect additional data and estimate parameters specifying the covariation, or else go ahead and fit our simple model recognising the biases introduced by the genotype-environment covariation. In these data we have no evidence for inequality of the total variances between zygosities and the simple model described in Section 3 may be fitted to the two types of twin jointly.

2.3 Inequality of the Variances between Races

Heterogeneity of the total variances between races could be due to inequalities of genetical or environmental components of variation. We saw in Section 2.5 that the mean score of Whites is higher than that of Blacks for all tests. Therefore, inequality of variances could be due to non-random sampling, genotype-environment interaction or to other genetical or environmental differences between the two populations.

From Table A41 we see that those tests which showed evidence for heterogeneity of total variances have a significant races effect. Therefore, at least some of the heterogeneity may be explained by

inequality of the total variances between the two races. We shall fit models separately to Whites and Blacks for all tests, not only those whose variances were heterogeneous. We can then compare estimates of the genetical and environmental components of variation in the two races.

2.4 Inequality of the Variances between Sexes

We have seen that the mean of males is higher than that of females for Spatial tests and that females score higher on average than males for Arithmetic and Spelling. We have discussed the possibility of sex-linkage or sex-limitation. In Section 4.7.1 we showed that sex-linkage may lead to inequality of the total variances so that males are more variable than females. From Table A41, we see no evidence for a sex effect for any of the Verbal-Arithmetic tests, but there are significant sex effects for several Spatial tests. This is the strongest evidence we have so far for sex-linkage for Spatial ability. However, this inequality of the total variances could be produced by different environmental components in the two sexes, by sex-limitation, by inadequate sampling or by genotype-environment interaction. We will fit models separately to males and females so that genetical and environmental components of variation may be compared in the two sexes.

3 FITTING MODELS TO THE VARIANCES

3.1 Model fitting Procedure

The total variances were partitioned into between and within pairs mean squares and the between pairs mean squares were corrected for the regression of test score on age. The corrected mean squares were given in Table A24.

The simple empirical model discussed in Section 3 will be fitted to the mean squares. If significant heritable variation is detected, we will then fit the theoretical model of Section 3 in order to determine the type of gene action producing variation among individuals in their test performance.

The expectations for the between and within pairs meansquares for monozygotic and dizygotic twins reared together were given in Table A28b. In order to determine the simplest model for explaining variation the adequacy of three models was tested in turn:

1. E_1
2. E_1E_2
3. G_1BE_1

These models were fitted to all groups of twins jointly, to each racial group separately and then to each race-sex group separately, using weighted least squares. As described in Section 3 and Appendix E, the variance of a meansquare is given by:

$$V^1 = \frac{2V^2}{N}$$

where V^{-1} is the variance of a mean square

V is the mean square

N is the number of degrees of freedom on which the mean square was based.

Thus, the weights form a diagonal matrix whose terms are the inverse of the variances of the appropriate mean squares. The off-diagonal elements are zero since the meansquares are independent. During weighted least squares, the variance-covariance matrix of the estimates is found as the inverse of $(\underset{\sim}{A}'\underset{\sim}{W}\underset{\sim}{A})$,

where $\underset{\sim}{A}$ is the model matrix

and $\underset{\sim}{W}$ is the diagonal matrix of weights.

An example of a variance-covariance matrix is given in Table A42 for the case of fitting the $G_1 B E_1$ model to the White males. The square roots of the diagonal terms are the standard errors of the estimates of G_1, B and E_1 shown in Table A46a and are used in testing the significance of the estimates. The two-tailed c-test is used, because, although the estimates are variance components and expected to be positive, if an inappropriate model is fitted estimates may turn out to be significantly negative.

When we fit models to race-sex groups separately, there are four different types of statistic (since we have only two types of relative). Therefore, we have only four degrees of freedom for estimating parameters and testing the adequacy of the model. The degrees of freedom for testing the adequacy of the model are given by (n-k), where n is the number of statistics and k is the number of parameters estimated from the data. Since the maximum number of parameters that we can estimate from

TABLE A42: VARIANCE-COVARIANCE MATRIX FROM FITTING THE $G_1 B E_1$ MODEL TO THE MEAN SQUARES OF THE ARITHMETIC TEST

	G_1	G_2	E_1
G_1	6576.1642	-4700.5598	-470.1793
G_2	-4700.5598	8836.0900	99.6615
E_1	-470.1793	99.6615	506.4501

NOTE: The Variance-Covariance Matrix is that for White Males

this data set is three (see Section 3), there is always at least one degree of freedom left for testing the adequacy of the model.

3.2 Models for the Eight Groups of Twins Jointly

The E_1, E_1E_2 and $G_1B\,E_1$ models were fitted in turn to the sixteen statistics provided by the eight groups of twins. The chisquare testing the adequacy of the E_1 model was highly significant for all tests. The estimates are not given since the model did not explain variation between individuals in performance. The failure of this model indicates the presence of significant between families variation or significant differences between the mean squares of MZ and DZ twins.

The E_1E_2 model

The E_1E_2 model tests the assumption that all between families variation can be explained by between families environmental effects (E_2) and all within families variation in terms of E_1. This model tests that the ratio σ_b^2/σ_w^2 is the same for all types of relative in the study (and that $\sigma^2_{bmz} = \sigma^2_{bdz}$ and $\sigma^2_{wmz} = \sigma^2_{wdz}$). If this model fails we assume that there is a heritable component of variation in ability between individuals since no adequate environmental model has been proposed which predicts that σ_b^2/σ_w^2 differs for different relatives. Several environmental models have been proposed. However, these may all be criticised on a number of grounds. Let us take for example, the model of Goldberger (personal communication). This model has no genetic components, but allows the environment to be more highly correlated in monozygotic twins. The model is specified:

Phenotype = $x_1 + x_2 + x_3$

where x_1 = an environmental component which is common to both members of a twin pair;

x_2 = an environmental component which is common to both members of an MZ pair, but independent for members of a DZ pair;

x_3 = an environmental component which is independent for members of a twin pair.

The x's are defined as mutually independent with variances V1, V2 and V3. Therefore, the total variance is V1 + V2 + V3. In terms of the mean squares, the model may be written:

	V1	V2	V3
MZ Between pairs	2	2	1
Within pairs	0	0	1
DZ Between pairs	2	1	1
Within pairs	0	1	1

It is obvious that this "environmental model" is merely a reformulation of the familiar $G_1B\ E_1$ model which we have discussed in Section 3. V3 is equivalent to our specific environmental component, E_1, and V1 is equivalent to our parameter B, where $B = G_2 + E_2$. When only data on twins reared together are available, the effects of genetical and environmental components operating between families cannot be separated. Therefore, their coefficients in the model are the same. From Table A28, we see that the coefficients of V2 are the same as those for G_1. G_1 has been re-defined as a variance component due to environmental effects common to members of an MZ pair, but different in DZ pairs, rather than

genetical effects common to members of an MZ pair and different in DZ pairs. Given that our genetical model and the "environmental" model are identical, how can we choose between the two formulations?

When we defined G_1 in Section 3, we had a definite mechanism of gene action in mind. Identical twins obtain identical genes from their parents; non-identical twins obtain a different selection of the parental genes. However, V2 is arbitrary. We have no model for the environment which enables us to suggest a mechanism for V2 differences. In this sense the genetical model is the better model since it is rooted in theory. It can be adequately tested in these data and further predictions based on it can be made and tested. The biggest weakness of the environmental model is that it has no predictive value. It is not clear how it could be extended to other types of relationship, whereas the genetical model can make predictions for any degree of relationship. Thus, the genetical model is the "strongest" model, since it can be falsified by testing the predictions it makes.

The weaknesses of Goldberger's "environmental" model are true for other environmental models proposed. The parameters are arbitrary with no clear mechanisms producing variation proposed. The models are specific to twins. They cannot be extended to other degrees of relationship and they cannot make further testable predictions. Thus, we conclude that no sensible environmental model apart from the E_1E_2 model has been formulated. Therefore, we shall test the E_1E_2 model and then move onto fit the simple, empirical genetical model, $G_1B\,E_1$.

The estimates and standard errors obtained by fitting the E_1E_2 model to all the data jointly are shown in Table A43 together with the corresponding chisquares for testing the adequacy of the model.

TABLE A43: THE E_1E_2 MODEL FOR THE EIGHT GROUPS OF TWINS

	E_1	se	E_2
Arithmetic	191.853	14.221***	475.642
Vocabulary	6.135	0.455***	4.335
Heim	89.632	6.644***	214.893
Spelling	50.071	3.712***	133.478
Calendar	21.258	1.576***	18.668
Identical pictures	100.165	7.425***	142.693
Cubes	41.470	3.074***	25.761
Surface Development	46.073	3.415***	68.916
Formboard	19.148	1.419***	16.317
Paper folding	8.945	0.663***	8.409
Object aperture	16.360	1.213***	13.512
Spatial ability	87.365	6.476***	204.863
Mazes completed	43.380	3.216***	37.233
Mazes errors	67.034	4.969***	32.389

se	x^2_{14}	P
43.910***	46.470	***
0.605***	21.461	5-10
19.967***	59.176	***
12.159***	72.846	***
2.357***	13.411	30-50
15.078***	40.164	***
3.845***	26.317	*
7.177***	89.150	***
2.087***	58.864	***
1.031***	17.622	20-30
1.753***	68.435	***
19.119***	53.728	***
4.748***	24.224	*
5.580***	173.066	***

We do not expect the model to fit those tests whose total variances are heterogeneous, since our models assume that the total variances are equal. Inequality of the total variances will cause failure of the model. The E_1E_2 model failed for all six tests whose total variances were heterogeneous, i.e. Spelling, Cubes, Surface development, Formboard, Object aperture and Spatial ability.

We can predict what we expect to happen in the model-fitting by looking at our analysis of the correlations. Of the eight tests whose total variances were homogeneous, no evidence of heterogeneity among the intraclass correlations could be found for Vocabulary, Calendar, Paper folding and Mazes Errors. The correlations for Arithmetic, Heim, Identical pictures and Mazes completed were heterogeneous. Therefore, we expect the E_1E_2 model to fit the first four tests, but not the latter four. A more powerful test for a difference between the correlations of monozygotic and dizygotic twins was made by fitting a linear model. The zygosities effect was significant for seven of the eight tests. Only the Calendar test showed no significant zygosities effect. Therefore, we expect that the simple, environmental model will fit the data for the Calendar test. From Table A43 we see that this happens. The simple E_1E_2 model is also adequate to explain variation for the other three tests whose correlations were homogeneous in the test for overall homogeneity. We predict that for those three tests, since a significant zygosities effect was found by fitting the linear model, that fitting the genetical model will produce a significant reduction in the residual chisquare and that the additional parameter will be significant.

The simple E_1E_2 model failed for the four tests whose correlations were heterogeneous and for which there was a highly significant zygosities effect, as we would expect. A more complex model is required to explain variation among individuals in performance on these tests. At the present time, no suitable, more complex two or three parameter environmental model has been proposed.

The $G_1B\ E_1$ model

In the absence of an environmental model predicting that σ_b^2/σ_w^2 differs for different relatives, the simple genetical model was fitted to all the data jointly. The results of fitting the $G_1B\ E_1$ model are shown in Table A44. The model fails for those six tests whose total variances were heterogeneous, as we predicted. i.e. None of our models will fit all the data jointly for these tests. From Table A41, we see significant heterogeneity between the races in their total variances for these six tests. Therefore, models must be fitted separately to Whites and Blacks. We expect that all models will fail when fitted to Whites and Blacks separately, over sexes, for Surface development, Object aperture and Spatial ability, since there was significant heterogeneity of total variances between sexes for these tests, (see Table A41). If this happens models will be fitted to each race-sex group separately.

Of the remaining eight tests the model fails only for Identical Pictures. This is surprising since the total variances for this test were homogeneous and there was no evidence for a races or sexes effect on the total variances. However, we recall that highly significant zygosities and races effects were detected when we fitted a linear model

TABLE A44:THE $G_1B\ E_1$ MODEL FOR THE EIGHT GROUPS OF TWINS

	G_1	se	B
Arithmetic	163.918	34.013 ***	382.400
Vocabulary	2.553	0.916 **	2.734
Heim	77.108	15.691 ***	168.450
Spelling	48.090	9.211 ***	104.101
Calendar	1.197	3.036 NS	17.875
Identical pictures	28.613	15.337 NS	125.054
Cubes	5.410	5.813 NS	22.456
Surface development	36.420	7.884 ***	47.457
Formboard	2.242	2.753 NS	14.901
Paper folding	2.455	1.336 NS	6.820
Object aperture	2.759	2.359 NS	11.923
Spatial ability	65.711	15.100 ***	166.788
Mazes completed	11.041	6.371 NS	30.739
Mazes errors	13.761	9.305 NS	23.117

se	E_1	se	χ^2_{13}	
49.482 ***	122.028	11.881 ***	9.569	70-80
0.821 ***	5.128	0.490 ***	11.868	50-70
22.509 ***	57.565	5.604 ***	15.895	20-30
13.541 ***	30.168	2.939 ***	42.733	***
2.950 ***	20.804	1.984 ***	13.200	30-50
17.765 ***	88.502	8.548 ***	38.540	***
5.160 ***	39.287	3.715 ***	25.506	*
8.773 ***	30.822	2.991 ***	54.357	***
2.648 ***	18.262	1.742 ***	56.877	***
1.309 ***	7.994	0.766 ***	13.898	30-50
2.253 ***	15.188	1.449 ***	66.840	***
21.560 ***	59.580	5.798 ***	28.314	**
6.124 ***	38.765	3.708 ***	21.108	5-10
7.832 **	62.097	10.650 ***	171.276	***

to the intraclass correlations and that the sexes effect was significant at the five per cent level. Thus, the relative magnitudes of the genetical and environmental components of variation must differ between groups without leading to any significant heterogeneity in the total variances i.e. $\sigma^2_T = G_1 + G_2 + E_1 + E_2$ is the same for all groups, but the G's and E's differ in their contribution to σ^2_T in the different groups. Therefore, models fitted to all the data jointly are not appropriate.

The $G_1B\ E_1$ model explains variation adequately for seven of the tests whose total variances were homogeneous. This model tests the assumption that the mean squares of monozygotic twins and dizygotic twins differ in a manner predictable on the basis of a simple genetical model. However, for Calendar and Paper folding, G_1 does not differ significantly from zero. Therefore, these parameters are redundant. We recall that the E_1E_2 model was adequate to explain variation in these tests. The difference between the chisquares of the $G_1B\ E_1$ and E_1E_2 models is a chisquare for one degree of freedom. This chisquare is not significant for Calendar and Paper folding and so fitting the $G_1B\ E_1$ model has not significantly reduced the residual chisquare. The E_1E_2 model is therefore, the most appropriate model for these tests.

In the case of Vocabulary, although the E_1E_2 model was adequate to explain variation, fitting the $G_1B\ E_1$ model significantly reduced the residual chisquare. Estimates of $G_1B\ E_1$ were all significantly greater than zero. This suggests that the $G_1B\ E_1$ model is the most appropriate model for this test. This model was also appropriate for Arithmetic, Heim and Mazes completed. The residual chisquare was non-significant and estimates of G_1B and E_1 were significantly greater than zero.

The two parameter E_1E_2 model did not explain variation in performance on these three tests.

Comparison with the Analysis of the Correlations

For the seven tests whose total variances are homogeneous and for which we have found an adequate model for all the data jointly, the results of the model fitting are broadly consistent with the results from the analysis of the correlations. The correlations of Arithmetic, Heim and Mazes completed were heterogeneous. These tests showed a highly significant zygosities effect when the linear model was fitted. The simple, genetical model was the only appropriate model for variation in performance on these tests, as we would have predicted. Conversely, the correlations for Calendar were homogeneous; there was no evidence for a zygosities effect. The E_1E_2 model was the only appropriate model. For Vocabulary, Paper Folding and Mazes errors, the results of the analysis of the correlations were ambiguous. Their correlations were homogeneous overall and yet a zygosities effect, significant at the five per cent level, could be detected. Model fitting has clarified our explanation of variation in performance on these tests. The simple environmental model is the most likely explanation of variation for Paper folding and the simple genetical model is the most likely for Vocabulary.

Analysis of correlations, when the total variances of the groups on which they are based are homogeneous, is valid and we expected the results obtained by analysing the correlations and variances to be consistent (see Section 4). However, for the remaining six tests whose

total variances are heterogeneous and for Identical pictures, the analysis of correlation coefficients is inappropriate and inconsistencies between the results of an analysis of correlations and of variances may be expected, when we fit models to races separately.

3.3 Models for Different Groups of Twins

Models for Whites and Blacks separately

The major factor contributing to heterogeneity of total variances is the effect of race (see Table A41). Since we are interested in comparing Whites and Blacks we shall fit models to races separately for all fourteen measures of ability. The E_1, E_1E_2 and $G_1B\ E_1$ models were fitted in turn to the mean squares for Whites and Blacks separately. The results of fitting the E_1E_2 model are given in Table A45, parts A and B. The results for the $G_1B\ E_1$ model are given in Table A46, parts A and B.

Models for Males and Females Separately

The effect of sex on the total variances was significant for Surface development, Object aperture and Spatial ability. Therefore, we expect all models to fail when fitted to Whites and Blacks over sexes. This happens for both Whites and Blacks for Object aperture, for Whites only for Surface development and Blacks only for Spatial ability. Therefore, we shall fit models for these three tests to each race-sex group separately. There are only four statistics for estimating

TABLE A45a: THE E_1E_2 MODEL FOR WHITES

	E_1	se	E_2
Arithmetic	203.044	18.125***	417.773
Vocabulary	6.418	0.573***	5.012
Heim	88.363	7.888***	217.140
Spelling	44.926	4.010***	117.192
Calendar	22.444	2.003***	18.262
Identical pictures	74.408	6.642***	146.909
Cubes	43.791	3.909***	31.135
Surface development	52.924	4.724***	87.999
Formboard	22.382	1.998***	21.733
Paper folding	8.910	0.795***	9.152
Object aperture	19.094	1.704***	16.439
Spatial ability	103.072	9.201***	217.502
Mazes completed	39.367	3.514***	41.332
Mazes errors	42.466	3.791***	19.139

se	χ^2_6	p
47.975***	28.831	***
0.799***	11.507	5-10
24.033***	52.160	***
12.827***	59.822	***
2.856***	8.615	10-20
17.030***	9.756	10-20
5.193***	9.056	10-20
10.649***	37.492	***
3.150***	10.264	10-20
1.297***	8.894	10-20
2.507***	24.532	***
24.837***	20.704	**
5.808***	10.121	10-20
4.124***	16.934	*

TABLE A45b: THE E_1E_2 MODEL FOR BLACKS

	E_1	se
Arithmetic	166.996	22.217***
Vocabulary	5.505	0.732***
Heim	92.451	12.230***
Spelling	61.500	8.182***
Calendar	18.624	2.478***
Identical pictures	157.376	20.937***
Cubes	36.314	4.831***
Surface development	30.854	4.105***
Formboard	11.965	1.592***
Paper folding	9.022	1.200***
Object aperture	10.288	1.369***
Spatial ability	52.478	6.982***
Mazes completed	52.296	6.957***
Mazes errors	121.606	16.178***

E_2	se	χ^2_6	p
609.049	96.224***	11.293	5 - 10
2.754	0.844**	3.354	70 - 80
209.750	35.859***	7.876	20 - 30
171.409	28.199***	9.777	10 - 20
19.554	4.172***	3.526	70 - 80
134.134	31.182***	3.704	70 - 80
13.215	4.958**	10.971	5 - 10
24.435	5.872	9.679	10 - 20
3.631	1.547**	9.204	10 - 20
6.690	1.659***	8.196	20 - 30
6.611	1.761***	20.987	**
174.891	27.977***	18.074	**
27.935	8.235***	7.155	30 - 50
64.195	19.054***	15.114	*

TABLE A46a: THE $G_1 B E_1$ MODEL FOR THE WHITES

	G_1	se	B
Arithmetic	171.876	41.899***	324.285
Vocabulary	3.133	1.169**	3.028
Heim	86.195	19.041***	167.738
Spelling	52.393	10.161***	87.365
Calendar	1.670	3.820NS	17.217
Identical pictures	35.783	14.303*	124.880
Cubes	8.991	7.466NS	26.256
Surface development	46.397	11.014***	60.311
Formboard	3.675	3.914NS	19.398
Paper folding	1.327	1.558NS	8.344
Object aperture	3.090	3.298NS	14.634
Spatial ability	73.299	20.530***	180.180
Mazes completed	6.394	6.902NS	37.543
Mazes errors	11.740	6.975NS	12.761

se	E_1	se	χ^2_5	ρ
55.279***	126.644	15.037***	2.716	70-80
1.055**	5.174	0.604***	3.582	50-70
27.001***	50.856	6.045***	10.475	5-10
14.427***	22.065	2.624***	8.631	10-20
3.594***	21.770	2.526***	8.372	10-20
19.129***	59.373	7.032***	2.548	70-80
6.800***	39.747	4.611***	7.744	10-20
12.649***	33.211	3.938***	13.485	*
3.882***	20.907	2.442***	8.835	10-20
1.582***	8.359	0.977***	7.807	10-20
3.159***	17.775	2.070***	23.660	***
28.386***	69.690	8.271***	5.631	30-50
7.068***	36.656	4.289***	9.384	5-10
5.799**	37.165	4.266***	11.534	

TABLE A46b: THE G_1BE_1 MODEL FOR BLACKS

	G_1	se		B
Arithmetic	138.420	56.327	*	542.582
Vocabulary	1.151	1.402	NS	2.049
Heim	57.521	29.300	*	171.188
Spelling	39.618	19.729	*	144.780
Calendar	-0.784	4.812	NS	20139
Identical pictures	18.351	41.379	NS	123.008
Cubes	4.180	8.910	NS	9.527
Surface development	11.178	8.447	NS	18.143
Formboard	-1.823	2.750	NS	4.794
Paper folding	6.464	2.539	*	1.787
Object aperture	1.279	2.642	NS	5.896
Spatial ability	38.611	17.715	*	143.589
Mazes completed	21.400	13.868	NS	15.663
Mazes errors	20.277	31.338	NS	48.598

se		E_1	se	χ^2_5	ρ
104.155	***	112.957	19.208***	2.121	80 - 90
1.224	NS	5.055	0.833***	2.562	70 - 80
40.767	***	71.055	12.054***	2.021	80 - 90
31.076	***	46.687	7.927***	6.278	20 - 30
5.098	***	18.882	3.162***	3.492	50 - 70
40.657	**	150.150	25.046***	3.368	50 - 70
7.393	NS	35.507	5.793***	11.008	5 - 10
8.059	*	26.186	4.380***	8.103	10 - 20
2.255	*	12.643	2.029***	8.874	10 - 20
2.150	NS	7.124	1.169***	4.140	50 - 70
2.424	*	9.748	1.616***	20.888	***
29.738	***	39.326	6.681***	12.902	*
12.125	NS	43.397	7.205***	5.597	30 - 50
27.330	NS	115.745	18.872***	14.698	*

parameters and testing the adequacy of the model. The $G_1B\ E_1$ model is the most complex that we will fit to the race-sex groups. This model has three parameters. Therefore, one degree of freedom remains for testing the adequacy of the model. This degree of freedom tests for effects leading to heterogeneity of the total variances. If the total variances are heterogeneous, this χ_1^2 will be significant and the model will fail. The sample size has been considerably reduced by subdividing the data into race-sex groups. Therefore, our tests of significance are now far less powerful.

The E_1, E_1E_2 and $G_1B\ E_1$ models were fitted in turn to the mean squares of each race-sex group for Surface development, Object aperture and Spatial ability. The estimates, standard errors and chisquares for testing the adequacy of the model for the E_1E_2 model and the $G_1B\ E_1$ model are shown in Table A47, parts A and B.

The Appropriate Models

Firstly, we shall compare the results obtained by fitting models jointly to all the data with those obtained by fitting models to Whites and Blacks separately. The E_1E_2 model explained variation for all the data jointly for Calendar and Paper folding. This model is also appropriate for Whites and Blacks separately. Fitting the three parameter genetical model does not significantly reduce the residual chisquare. The $G_1B\ E_1$ was the most appropriate for all the data jointly for Arithmetic, Vocabulary, Heim and Mazes completed. For both races on Arithmetic and Heim and for Whites on Vocabulary, the $G_1B\ E_1$ model was again the most likely model for the observed variation. However, for

TABLE A47a: THE E_1E_2 MODEL FOR SEXES SEPARATELY FOR THREE TESTS

		E_1	se	E_2	se
White males	Surface development	50.056	6.812***	133.956	22.309 *
	Object aperture	24.324	3.310***	24.544	5.353 *
	Spatial ability	105.319	14.332***	248.188	42.331 *
White females	Surface development	55.091	6.515***	53.603	10.265 *
	Object aperture	15.143	1.780***	10.393	2.274 *
	Spatial ability	101.374	11.989***	194.550	30.021 *
Black males	Surface development	34.368	8.293***	42.228	15.841 *
	Object aperture	13.206	3.203***	12.256	5.126 *
	Spatial ability	71.015	17.224***	241.029	71.919 *
Black females	Surface development	29.342	4.669***	17.371	5.730 *
	Object aperture	9.032	1.425***	4.397	1.516 *
	Spatial ability	44.500	7.080***	148.717	28.142 *

	x^2_2	p
•••	10.095	••
•••	0.808	50 - 70
•••	3.255	10 - 20
•••	13.602	••
•••	1.278	50 - 70
•••	16.545	•••
••	0.068	95 - 98
•	4.927	5 - 10
•••	2.513	20 - 30
••	4.650	5 - 10
••	1.809	30 - 50
•••	10.272	••

TABLE A47b: THE G_1BE_1 MODEL FOR SEXES SEPARATELY FOR THREE TESTS

Group	Test	G_1	se		B	se		E_1	se		χ^2_1	
White males	Surface development	44.144	17.472	*	105.871	25.235	***	33.186	5.764	***	0.086	70 - 80
	Object aperture	4.756	6.698	NS	21.830	6.834	**	22.375	3.827	***	0.245	50 - 70
	Spatial ability	50.128	32.459	NS	219.169	47.598	***	83.303	14.787	***	0.190	70 - 80
White females	Surface development	45.703	13.697	***	29.317	12.950	*	33.510	5.424	***	0.008	90 - 95
	Object aperture	2.778	3.346	NS	8.774	2.934	**	13.918	2.200	***	0.567	30 - 50
	Spatial ability	90.168	25.954	***	154.722	33.650	***	55.965	9.102	***	1.405	20 - 30
Black males	Surface development	-3.963	16.358	NS	44.772	18.970	*	35.790	10.651	***	0.011	90 - 95
	Object aperture	5.421	4.623	NS	2.345	4.175	NS	6.345	1.891	***	1.377	20 - 30
	Spatial ability	46.491	44.838	NS	203.526	77.492	**	56.232	16.928	***	1.158	20 - 30
Black females	Surface development	16.080	9.387	NS	9.347	8.305	NS	21.864	4.436	***	0.905	30 - 50
	Object aperture	0.666	2.580	NS	3.923	1.801	NS	8.814	1.719	***	1.738	10 - 20
	Spatial ability	37.135	18.282	*	118.502	29.660	***	31.538	6.501	***	5.352	*

Blacks on Vocabulary and for both races on Mazes completed, the E_1E_2 model was adequate to explain variation and fitting the $G_1B\ E_1$ model did not significantly reduce the residual chisquare. This is not consistent with the findings from fitting models to the races jointly.

We have not yet considered models for Spelling, Cubes, Surface development, Formboard, Object aperture, Spatial ability, or Identical pictures since it was inappropriate to fit models to all the data jointly for these tests. We can predict which models we expect to explain the variation in performance in these tests from the analysis of the correlations. There was no significant effect of zygosity on the correlations detectable by fitting the linear model for Cubes, Object aperture, or Formboard. Therefore, we expect that the E_1E_2 model will explain variation in test performance. We see that the E_1E_2 model is adequate to explain variation and that fitting the $G_1B\ E_1$ model does not significantly reduce the residual chisquare for these tests.

The effect of zygosity on the correlations was significant for the remaining four tests, i.e. Surface development, Identical pictures, Spatial ability and Spelling. Therefore, we expect that the genetical model will be required to explain variation in test performance. For Spelling the genetical model is clearly required to explain variation in test performance. However, for the other three tests the results are ambiguous. The E_1E_2 model fits for Identical pictures. However, a significant reduction in residual chisquare is obtained by fitting the $G_1B\ E_1$ model in Whites, but not in Blacks. Similarly for Surface Development and Spatial ability, the genetical model is the most appropriate for some race-sex groups and the environmental model is adequate for others.

The results, so far, appear to display little regularity. All tests are measures of ability. We recall that a general factor of ability accounted for approximately fifty per cent of the variation (Section 2.1). Therefore, we might expect that either the genetical or the environmental model would be appropriate for all the tests. Given a substantial general factor, if one test displays heritable variation, then we expect variation in performance on all tests to be in some part heritable. The importance of genetical and environmental components in producing variation could, of course, differ from test to test, since there is a lot of specific variation. We have found that the environmental model appears to account for variation in some tests and the genetical model for variation in others. One explanation could be that the heritable component was smaller for some tests that others and, therefore, more difficult to detect.

We may now ask whether tests, having the same model as the most likely explanation of the variation, have anything in common. In Section 2.1 we showed that the tests loaded on two factors, one Verbal-Arithmetic and the other Spatial. This test structure may suggest an explanation of the results of the model fitting. A summary of the results of the model fitting is given in Table A48. This summary reveals some consistency in the results. The $G_1 B E_1$ model is required to explain variation for the tests of the Verbal-Arithmetic factor. However, for the tests of the Spatial factor, either the $E_1 E_2$ model is adequate to explain variation or else the $G_1 B E_1$ model is required for some race-sex groups and not others. Thus, these data are consistent with a simple, genetical model for the Verbal tests. There is no convincing evidence

TABLE A48: SUMMARY OF MODEL FITTING TO THE MEANSQUARES FROM THE GEORGIA TWIN STUDY

TEST	All data jointly	Races separately	Races and sexes separately
Arithmetic	G	G	
Vocabulary	G	G*	
Heim	G	G	
Spelling	-	G	
Calendar	E	E	
Identical pictures	-	G*	
Cubes	-	E	
Surface development	-	-	G*
Formboard	-	E	
Paper folding	E	E	
Object aperture	-	-	E
Spatial ability	-	-	G*
Mazes completed	G	E	
Mazes errors	E	E	

KEY

E — The E_1E_2 model was adequate for all groups.

G — The residual χ^2 from fitting the $G_1B\ E_1$ model was significantly less than the residual χ^2 from fitting the E_1E_2 model i.e. the $G_1B\ E_1$ model was the most appropriate model

G* — The $G_1B\ E_1$ model was the appropriate model for some race-sex groups, but was not needed to explain variation in test performance for other groups i.e. results somewhat ambiguous.

for a heritable component in the case of Spatial tests. The analysis of the correlations gave a broadly similar result. The effect of zygosity on the correlations was significant for all Verbal tests, but for only six of the Spatial tests (see Table A36). Four tests only showed highly significant zygosities effects and these are the four tests where a $G_1 B E_1$ model is required to explain variation for one or more of the race-sex groups. Paper folding showed a zygosities effect significant at the five per cent level. However, the $E_1 E_2$ model was adequate to explain variation for this test.

It is surprising that there are no greater inconsistencies between the analysis of the correlations and of the variances for those tests whose total variances were not homogeneous. However, this is probably because so far, we have only discussed the detection of effects and not their estimation. Comparison of estimates of genetical and environmental effects from the correlations and the mean squares might reveal bigger discrepancies between the two approaches.

An estimate of the biased broad heritability (Bh_b^2), inflated by any E_2, is given by the intraclass correlations for monozygotic twins reared together. We saw that generally this was higher for Verbal-Arithmetic than for Spatial tests. It was also higher for those tests showing a demonstrably significant effect of zygosity on the correlations. This leads us to suggest that the power of the test for detecting a heritable component of variation was less for tests with a low heritability. We showed that the probability of detecting a heritable component of variation for tests whose monozygotic twin intraclass correlations were as low as those for some of the Spatial tests was very small. Therefore, even if genetical effects produce variation in Spatial ability, we would

have been unlikely to detect this is most cases. In view of this, it seems that we should have expected the E_1E_2 model to be adequate to explain variation in Spatial ability measured by most of these tests, given the sample sizes employed.

3.4 Heritability of these measures of Ability

The broad heritability of a trait is that proportion of the total variation due to all genetical causes. In these data, it is possible to calculate

$$\hat{Bh}_b^2 = \frac{G_1 + B}{G_1 + B + E_1}$$

which is a biased estimate of the true broad heritability inflated by common environmental effects.

This biased estimate of heritability was calculated for each of the fourteen tests, using the estimates of G_1B and E_1 from Table A46, parts A and B and from Table A47a for Surface development, Object aperture and Spatial ability. The variance of $\hat{Bh}_b^2$ and hence its standard error, may be derived using the formula for the variance of a ratio given by Kempthorne (1957). If a/b is a ratio, then its variance is given by the formula:

$$V(a/b) = (a/b)^2 \left[V(a)/a^2 + V(b)/b^2 - 2W(a.b) / ab\right]$$

where $V(a/b)$ is the variance of the ratio

$V(a)$ is the variance of the numerator

$V(b)$ is the variance of the denominator

$W(a.b)$ is the covariance between the numerator and denominator.

When a/b is the biased broad heritability, then

$$a = G_1 + B$$

and $$b = G_1 + B + E_1$$

and it turns out that:

$$V(a/b)=V(Bh^2_b)=Bh^2_b\left[\frac{(V_{G_1}+V_B+2W_{G_1B})}{(G_1+B)^2}+\frac{(V_{G_1}+V_B+V_{E_1}+2W_{G_1B}+2W_{G_1E_1}+2W_{BE_1})}{(G_1+B+E_1)^2}\right.$$

$$\left.-2\,\frac{(V_{G_1}+V_B+2W_{G_1B}+W_{G_1E_1}+W_{BE_1})}{(G_1+B)(G_1+B+E_1)}\right]$$

The variances of G_1, B and E_1 and the covariances between them are the elements of the variance-covariance matrix calculated during the least squares procedure (i.e. of $(A'WA)^{-1}$). The variance of the biased broad heritability is found by evaluating the expression given above, using these values.

Estimates of $\hat{Bh}^2_b$ and their standard errors are given in Table A49 for each test.

In general, the biased estimate of heritability for Verbal-Arithmetic tests is higher than that for Spatial tests, which confirms that heritable variation will be harder to detect for Spatial than for Verbal-Arithmetic tests. Among the Spatial tests, those with the highest $\hat{Bh}^2_b$ were the tests where a heritable component was required to explain variation in test performance.

The heritabilities of Whites and Blacks were compared using a two-tailed c-test for the tests in Part A of Table A49. The values of $\hat{Bh}^2_b$ were sigmificantly different at the five per cent level for two of the eleven tests only i.e. Identical pictures and Formboard. The biased

TABLE A49: ESTIMATES OF THE BIASED BROAD HERITABILITY

PART A

	WHITE		BLACK			
	Bh^2_b	se	Bh^2_b	se	C	ρ
Spelling	0.864	0.019	0.798	0.040	1.500	NS
Heim	0.833	0.023	0.763	0.047	1.346	NS
Arithmetic	0.797	0.028	0.854	0.030	-1.390	NS
Vocabulary	0.544	0.056	0.388	0.101	1.357	NS
Identical pictures	0.730	0.036	0.485	0.090	2.526	*
Mazes completed	0.545	0.056	0.461	0.092	0.778	NS
Paper folding	0.536	0.057	0.537	0.078	-0.010	NS
Formboard	0.525	0.058	0.190	0.117	2.557	*
Cubes	0.470	0.063	0.279	0.111	1.492	NS
Calendar	0.465	0.064	0.506	0.087	0.380	NS
Mazes errors	0.397	0.069	0.373	0.102	0.195	NS

PART B

	WHITE				BLACK			
	Male		Female		Male		Female	
	Bh^2_b	se	Bh^2_b	se	Bh^2_b	se	Bh^2_b	se
Surface development	0.819	0.037	0.691	0.055	0.533	0.150	0.538	0.099
Spatial ability	0.760	0.048	0.814	0.034	0.816	0.067	0.832	0.041
Object aperture	0.543	0.083	0.455	0.088	0.606	0.131	0.342	0.123

PART C

	WHITE		BLACK			
	Bh^2_b	se	Bh^2_b	se	c	
Surface development	0.755	0.047	0.536	0.127	1.622	NS
Spatial ability	0.787	0.042	0.824	0.056	0.529	NS
Object aperture	0.499	0.086	0.474	0.127	0.163	NS

heritabilities of males and females were compared within each race for the tests in Part B of Table A49. The estimates of $\hat{Bh}_b^2$ for males and females were not significantly different in Whites or Blacks for any of the three tests. Therefore, the estimates of $\hat{Bh}_b^2$ for males and females were combined and the pooled standard errors calculated. These combined estimates are shown in Part C of Table A49. The c-test shows that the estimates of $\hat{Bh}_b^2$ for Whites and Blacks are not significantly different for any of the three tests.

Thus, significant differences between estimates of $\hat{Bh}_b^2$ have been found between Whites and Blacks only for Identical pictures and Formboard. The source of this difference in $\hat{Bh}_b^2$ will be investigated in a later section.

3.5 A Theoretical Genetical Model

The E_1E_2 and $G_1B\,E_1$ models yield estimates which may fruitfully be compared between tests and between race and sex groups. However, before making comparisons we shall consider the theoretical model described in Section 3. We have seen that a likely explanation for failure detect a heritable component of variation for the Spatial tests is the low power of the test. In comparing the adequacy of the E_1E_2 and $G_1B\,E_1$ models we are comparing a two parameter model with a three parameter model. The two parameter model is the simplest model and should be accepted as the most likely cause of variation if chisquare is not significant. This model tests the assumption that all variation between individuals may be explained by environmental differences between them. However, we have not tested any two parameter genetical model of

variation. Therefore, we cannot conclude that there is no heritable component of variation for the Spatial tests, but only that the data are consistent with a purely environmental interpretation.

The $G_1 B E_1$ model is the simplest empirical model that we can fit to the data. However, the genetical theory discussed in Section 3 shows that under some circumstances we can simplify this model still further and formulate a two parameter genetical model. If we assume that there are no between families environmental effects (i.e. E_2) and that $G_1 = G_2 = \frac{1}{4} D_R$, then we can fit an $E_1 D_R$ model to these data. This model tests the hypothesis that all variation may be explained in terms of additive genetic variation and specific environmental variation.

The expectations for this two parameter model for the mean squares of monozygotic and dizygotic twins reared together were given in Table A32. The assumptions made in formulating this model are:

1. No E_1
2. No dominance
3. No assortative mating
4. No non-allelic interactions
5. No genotype-environmental interactions
6. No genotype-environmental covariation
7. Equality of E_1's for all types of relatives

The latter three assumptions were also made in formulating the $G_1 B E_1$ model and may be realistic for certain traits. Since failure of these assumptions may lead to heterogeneity of the total variances, we expect to detect their failure by failure of the $G_1 B E_1$ model. The first four

assumptions were not made in formulating the simple empirical model, although an independent estimate of E_2 could not be obtained, and only $\hat{B} = G_2 + E_2$ could be estimated.

Although $G = G_1 + G_2$ in our simple model, there was no constraint that $G_1 = G_2$. We showed in Section 3 that in the absence of dominance and assortative mating we expect $G_1 = G_2$. Therefore, we had a test for dominance or assortative mating since if $G_1 > G_2$ dominance has been demonstrated and if $G_2 > G_1$ there is assortative mating. Of course with this data set we can only test whether $B > G_1$ and, therefore, the effects of assortative mating and E_2 are inevitably confounded. When we fit the simple E_1D_R model failure of any of the assumptions stated above would produce failure of the model. If this model fails we may have little idea which of the assumptions has failed unless there is evidence from another source. One of more of the assumptions we have made to arrive at the simplicity of the additive genetical model are likely to fail for any complex trait, given enough pairs of twins to provide a powerful test.

We will fit the simple additive genetical model to these data, so that the E_1E_2 and E_1D_R models may be compared. However, we must stress that this model is an unrealistic one for any biologically important trait, although perhaps no more unrealistic than the E_1E_2 model.

Fitting the E_1D_R model

The model for monozygotic and dizygotic twins reared together, specified in Section 3, was fitted to the data by weighted least squares where the weights are the same as those for the $G_1B\,E_1$ model. The model was fitted

to all the data jointly, to each race separately and then to each race-sex group separately for Surface development, Object aperture and Spatial ability. The estimates, their standard errors and chisquares for testing the adequacy of the model are given in Tables A50; A51, Parts a and b and A52 Parts a, b, c and d.

An estimate of heritability may be obtained:

$$\hat{h}_b^2 = \hat{h}_n^2 \frac{\frac{1}{2}D_R}{\frac{1}{2}D_R + E_1}$$

where h_n^2 is the proportion of the total variation produced by additive genetical variation. Since our simple genetical model assumes no non-additive effects, the narrow heritability equals the broad heritability. The variance of the heritability can be found as the variance of a ratio as previously described for $\hat{Bh}_b^2$ (see Section 5.3.4).. It turns out that the variance of the heritability, after simplification, is given by:

$$V_{(h^2)} = V\left(\tfrac{1}{2} D_R / \tfrac{1}{2} D_R + E_1\right)$$

$$= \frac{E_1^2 V_{(D_R)} + D_R^2 V_{(E_1)} - 2D_R E_1 W(D_R E_1)}{4\left(\tfrac{1}{2} D_R + E_1\right)^4}$$

The variances of D_R and E_1 and their covariance are the elements of the variance-covariance matrix calculated during the model fitting procedure. Thus the variance of the heritability was found by evaluating the expression for $V(h^2)$ given above. Estimates of heritability and their standard errors are given in the final columns of Tables A50; A51, parts a and b and A52 parts a, b, c and d.

TABLE A50: ESTIMATES OF E_1 and D_R FOR ALL THE DATA JOINTLY

Test	E_1	se	D_R	se
Arithmetic	117.412	11.303 ***	1055.030	82.350 ***
Vocabulary	5.113	0.465 ***	10.592	1.348 ***
Heim	55.591	5.346 ***	475.286	37.467 ***
Spelling	29.214	2.817 ***	292.909	22.405 ***
Calendar	19.505	1.773 ***	39 .430	5.079 ***
Identical pictures	83.153	7.776 ***	306.784	30.543 ***
Cubes	37.591	3.370 ***	58.375	8.645 ***
Surface development	30.397	2.886 ***	166.901	14.633 ***
Formboard	17.219	1.566 ***	35.507	4.529 ***
Paper folding	7.681	0.705 ***	18.803	2.211 ***
Object aperture	14.352	1.307 ***	30.563	3.837 ***
Spatial ability	57.059	5.477 ***	449.821	36.065 ***
Mazes completed	37.104	3.393 ***	85.664	10.351 ***
Mazes errors	61.232	5.426 ***	75.049	12.774 ***

χ^2_{14}	p	Bh^2_b	se
16.530	20-30	0.818	0.021
11.869	50-70	0.509	0.047
18.084	20-30	0.810	0.021
49.423	***	0.834	0.019
18.730	10-20	0.503	0.048
45.887	***	0.649	0.037
29.137	*	0.437	0.052
56.012	***	0.733	0.029
60.997	***	0.508	0.048
16.594	20-30	0.550	0.045
74.169	***	0.516	0.047
37.341	***	0.798	0.023
24.048	*	0.536	0.046
171.762	***	0.380	0.055

TABLE A51: PART a - ESTIMATE OF E_1 and D_R FOR WHITES

Test	E_1	se	D_R
Arithmetic	122.527	14.338 ***	970.674
Vocabulary	5.182	0.577 ***	12.313
Heim	49.325	5.804 ***	491.903
Spelling	21.599	2.550 ***	271.323
Calendar	20.397	2.249 ***	39.496
Identical pictures	56.022	6.490 ***	312.822
Cubes	37.917	4.179 ***	73.499
Surface development	32.760	3.813 ***	211.807
Formboard	19.708	2.198 ***	47.211
Paper folding	7.814	0.874 ***	19.862
Object aperture	16.739	1.857 ***	36.821
Spatial ability	66.397	7.756 ***	494.602
Mazes completed	34.217	3.838 ***	90.314
Mazes errors	37.035	4.002 ***	49.242

se	χ^2_6	p	h^2_b	se
93.064 ***	4.703	50-70	0.798	0.027
1.767 ***	3.583	70-80	0.543	0.055
45.036 ***	10.945	5-10	0.833	0.023
23.913 ***	9.747	10-20	0.863	0.019
6.261 ***	11.975	5-10	0.492	0.059
32.715 ***	8.429	20-30	0.736	0.035
11.647 ***	9.885	10-20	0.492	0.059
21.312 ***	14.526	*	0.764	0.032
6.747 ***	11.356	5-10	0.545	0.054
2.763 ***	11.632	5-10	0.560	0.053
5.480 ***	28.345	***	0.524	0.056
48.080 ***	11.901	5-10	0.788	0.029
12.352 ***	15.234	*	0.569	0.052
9.624 ***	11.423	10-20	0.399	0.065

TABLE A51:PART b - ESTIMATES OF E_1 and D_R FOR BLACKS

Test	E_1	se	D_R
Arithmetic	107.999	18.260 ***	1241.656
Vocabulary	4.963	0.778 ***	6.555
Heim	68.107	11.378 ***	441.570
Spelling	44.552	7.476 ***	350.231
Calendar	17.647	2.829 ***	38.850
Identical pictures	141.494	22.646 ***	294.507
Cubes	35.305	5.403 ***	27.421
Surface development	25.596	4.115 ***	59.488
Formboard	11.867	1.808 ***	7.351
Paper folding	7.413	1.189 ***	16.232
Object aperture	9.262	1.466 ***	15.210
Spatial ability	37.848	6.357 ***	338.458
Mazes completed	43.981	6.961 ***	73.143
Mazes errors	113.540	17.691 ***	140.523

se		χ^2_6	p	Bh^2_b	se
169.303	***	7.803	20-30	0.852	0.030
1.908	***	2.692	80-90	0.398	0.096
67.320	***	3.787	70-80	0.764	0.046
51.145	***	9.866	10-20	0.797	0.040
8.630	***	6.086	30-50	0.524	0.082
67.160	***	4.946	50-70	0.510	0.084
11.285	*	11.036	5-10	0.280	0.105
12.886	***	8.649	10-20	0.538	0.080
3.567	*	10.834	5-10	0.237	0.107
3.614	***	3.738	70-80	0.523	0.082
3.914	***	23.829	***	0.451	0.090
48.219	***	17.063	**	0.817	0.036
18.700	***	5.533	30-50	0.454	0.090
42.561	***	14.837	*	0.382	0.097

TABLE A52: ESTIMATES OF E_1 and D_R FOR SEXES SEPARATELY

PART A WHITE MALES	E_1	se	D_R
Surface development	32.100	5.515 ***	289.400
Object aperture	21.070	3.480 ***	54.950
Spatial ability	80.290	13.705 ***	523.300
PART B WHITE FEMALES			
Surface development	34.500	5.404 ***	150.200
Object aperture	13.290	1.982 ***	24.000
Spatial ability	54.080	8.693 ***	480.100
PART C BLACK MALES			
Surface development	32.730	9.420 ***	84.040
Object aperture	9.985	2.910 ***	32.68
Spatial ability	54.070	16.099 ***	462.600
PART D BLACK FEMALES			
Surface development	22.560	4.368 ***	49.680
Object aperture	8.585	1.614 ***	9.438
Spatial ability	30.460	6.227 ***	291.100

se		X^2_2	p	Bh^2_b
41.525	***	2.266	30-50	0.818
11.434	***	2.023	30-50	0.566
13.705	***	4.968	5-10	0.765
22.467	***	0.488	70-80	0.685
5.290	***	1.506	30-50	0.474
59.245	***	2.808	20-30	0.816
32.163	**	1.846	30-50	0.562
11.284	**	3.734	10-20	0.621
124.442	***	2.641	20-30	0.811
13.416	***	1.107	50-70	0.524
3.674	**	2.039	30-50	0.355
48.484	***	8.188	•	0.585

Comments on the E_1D_R Model

For the four tests of the Verbal-Arithmetic factor, the simple environmental model was not adequate to explain variation between individuals in test performance. However, we see that the simple additive genetical model is adequate to explain variation for these tests. Therefore, we have no evidence that any of the assumptions on which this simple model is based have failed. Variation may be explained by differences in environment specific to individuals and by the effect of genes with additive properties. No common environmental effects, dominance or more complicated gene action is necessary for explaining variation. However, failure to detect more complex effects may not be because they are absent but because the power of our test for detecting them is too low. For example, Eaves (1972) has shown that experimental sizes much larger than this are required for the detection of quite substantial amounts of dominance variation.

The simple genetical model was adequate to explain variation for most tests of the Spatial factor. Mazes completed was an exception. The $G_1B\ E_1$ model fitted adequately all the data jointly. However, for each race separately the E_1E_2 model was adequate to explain variation. The E_1D_R model fails for all the data jointly and for Whites, although it fits for Blacks. If we consider races separately, the E_1E_2 model is more appropriate for Mazes completed than the E_1D_R model. However, the total variances of this test were homogeneous, and the $G_1B\ E_1$ model was adequate to explain variation for both races jointly. In this case, we conclude that the simple additive genetical model is not adequate to explain variation and that one of the assumptions underlying the E_1D_R

model has failed. In this case, common environmental effects, dominance, assortative mating or more complex gene action might be required to explain variation (although dominance is unlikely since $\hat{B} > \hat{G}_1$).

The E_1E_2 model was adequate to explain variation for most of the other Spatial tests and the E_1D_R model fitted all of them. Thus, for the Spatial tests there are problems in discriminating between simple genetical and simple environmental models, since both fit the data. This suggests that our experimental size is too small, since we should be able to discriminate between the two models given data on a sufficient number of pairs of twins. This problem did not arise for the Verbal-Arithmetic tests since $\hat{Bh}_b^2$ was higher overall for these tests and the power of the test for a genetical component was higher.

3.6 Comparison of the Estimates between Blacks and Whites

There is heterogeneity of total variances between races for six of our fourteen tests (Table A41). One or more components of variation must differ between the races for these tests. However, components of variation may differ between races when the total variances are homogeneous, since even with σ^2_T the same for both races, the relative magnitudes of the genetical and environmental components may differ. Estimates of the components of variance may be compared by testing whether the difference between the estimates in the two races is significantly greater than zero, using a c-test, where c is given by the difference between the estimates divided by the pooled error. We will attempt to explain differences between the races in variance in terms of differences in specific genetical or environmental components of variance.

Tests of the Spatial Factor

The total variances of five of the ten tests loading on the Spatial factor were heterogeneous. For these tests there was a significant positive effect of race on variance (Table A41) i.e. Whites were significantly more variable than Blacks in performance. Of the remaining five tests, the race effect was positive for four. The race effect for Identical pictures was non-significantly negative. We will attempt to explain the greater variability in Whites for all tests except Identical pictures in terms of the variance components.

First we shall compare estimates of $\hat{E}_1$ obtained by fitting the E_1D_R model to Whites and Blacks separately, using the c test. Estimates of $\hat{E}_1$ from the E_1D_R model were used since this was the only model adequately explaining variation for all the Spatial tests. However, the comparison of estimates from the E_1E_2 and $G_1B\,E_1$ models was also made and yielded similar results.

Estimates of $\hat{E}_1$ were larger in Whites than Blacks for all tests except Identical pictures and Mazes completed and estimates were significantly greater for Formboard, Object aperture and Spatial ability. An increased E_1 contributes to the increased variability of Whites in these tests. The estimate of $\hat{E}_1$ for Whites was not significantly less than that for Blacks for Mazes completed. However, the difference between the races was significant for Identical pictures. We recall that Whites were non-significantly less variable than Blacks for performance on this test. It appears that the smaller $\hat{E}_1$ in Whites must be contributing to this difference.

Similarly, we compared estimates of $\hat{E}_2$ obtained from fitting the E_1E_2 model to these data, although this model was not adequate to explain variation for all tests. Estimates of $\hat{E}_2$ were significantly greater in Whites for Cubes, Surface development, Formboard and Spatial ability and were greater in Whites for all the other tests of the Spatial factor, except Calendar. For Calendar, $\hat{E}_2$ was non-significantly smaller in Whites. If the best explanation of the data were the E_1E_2 model, E_2 might contribute to the greater variability in Whites. However, if the genetical model was more appropriate and we fitted the E_1E_2 model, then $\hat{E}_2$ would be inflated by the genetical component of variation.

We may generalise that E_1 but not necessarily E_2 contributes to the greater variability among Whites in performance on Spatial tests although Calendar, Identical pictures and Mazes completed have provided exceptions. $\hat{E}_2$ is not smaller in Whites for Identical pictures and for this test the reduced variability is due to E_1 only and not to E_2.

We will consider the relative magnitudes of the genetical parameters obtained by fitting the $G_1B\,E_1$ model. It turns out that $\hat{G}_1$ is non-significantly larger in Whites for all tests except Paper folding and Mazes completed. Estimates of $\hat{B}$ are larger in Whites than Blacks for all tests except Calendar. The comparison is significant for Surface development, Formboard, Paper folding and Object aperture.

Comparing estimates of $\hat{D}_R$ in the two races, we find that $\hat{D}_R$ is greater in Whites in all cases, and significantly greater for Cubes, Surface development, Formboard, Object aperture and Spatial ability i.e. all the tests where there was a significant positive effect of race on the total variances.

In summary, we can consider three types of Test: (A) those tests where the races effect was significant and positive, (B) those tests where the races effect was positive but not significant and (C) Identical pictures where the races effect was negative. The results are summarised in this format in Table A53. In this form, the results are surprisingly consistent. All components of variation are greater in Whites for the tests in Group A and contribute to the significant effect of race on the total variance. We cannot attribute the increased variation in Whites to any specific component of variation.

The difference between the components in Whites and Blacks is not significant for the tests of group B whose total variances are homogeneous, although in theory it would be possible for one component to be significantly greater in Whites and another to be significantly greater in Blacks and for the total variances to remain homogeneous.

For identical pictures the total variances of Whites were less than the total variances of Blacks. This may be explained by the significantly smaller $\hat{E}_1$ in Whites. The other components were non-significantly larger in Whites. Although $\hat{E}_1$ was significantly smaller in Whites, the other components were larger in Whites than Blacks and the total variances were homogeneous. We recall that our models failed when fitted to the races jointly although the total variances were homogeneous and we predicted that genetical or environmental components of variation must differ between the races. We can now confirm this prediction and attribute the failure of the models when fitted to the races jointly to the differences in the size of the E_1 component between the races.

TABLE A53: SUMMARY OF COMPARISON OF ESTIMATES BETWEEN BLACKS AND WHITES FOR SPATIAL TESTS

A. Races effect positive and significant

	E_1	E_2	G_1	B	D_R
Cubes	+0.38 NS	+2.50 *	+0.41 NS	+1.67 NS	+2.84 **
Surface development	+1.28 NS	+5.23 ***	+2.54 *	+2.81 **	+6.12 ***
Formboard	+2.76 **	+5.16 ***	+1.15 NS	+3.25 **	+5.22 ***
Object aperture	+3.16 **	+3.37 ***	+0.43 NS	+2.19 *	+3.21 **
Spatial ability	+2.85 **	+1.14 NS	+1.24 NS	+0.89 NS	+2.30 *

B. Races effect positive and non-significant

	E_1	E_2	G_1	B	D_R
Calendar	+0.76 NS	-0.26 NS	+0.40 NS	-0.47 NS	+0.06 NS
Paper folding	+0.27 NS	+1.17 NS	-1.72 NS	+2.46 *	+0.80 NS
Mazes completed	-1.23 NS	+1.33 NS	-0.97 NS	+1.56 NS	+0.77 NS

C. Races effect negative and non-signficant

Identical pictures	-3.63 ***	+0.36 NS	+0.40 NS	+0.04 NS	+0.25 NS
Mazes errors	-4.22 ***	+2.31 *	-0.27 NS	-1.28 NS	-2.10 *

NOTE:

1. The values given in the body of the table a values of c for comparing the components between the races

2. When c is positive the component is larger in Whites than Blacks
   When c is negative the component is larger in Blacks than Whites

Non-genetical variation specific to individuals (i.e. E_1) may be produced by accidents of development and pre-natal influences or post-natal experiences specific to individuals. We expect these environmental factors to produce comparable variation among individuals in performance on all Spatial tasks. Thus for $\hat{E}_1$ to be smaller in Whites for Identical pictures, but larger for the other Spatial tests seems anomalous. We recall that originally Identical pictures and the Mazes tests were chosen to represent a factor of Perceptual speed, rather than Spatial ability. In these data, we only found one Spatial factor, possibly because of under representation of tests of Perceptual speed in the test battery. From Table A53, we see that $\hat{E}_1$ is non-significantly smaller for the Mazes test. It is possible that the Identical pictures and Mazes tests are measuring a different ability and the specific environmental effects contributing to variation in performance on these tests are different to those affecting variation for other Spatial tests. Therefore, $\hat{E}_1$ is smaller in Whites than Blacks for tests of Perceptual speed, but greater in Whites for other measures of Spatial ability.

However, we recall that there was a high level of test specific variation for Identical Pictures and Mazes errors. We suggested that this might be due to low test reliability, which would inflate $\hat{E}_1$. It is possible that the Identical pictures tests is more reliable in one race than the other.

Tests of the Verbal-Arithmetic Factor

The chisquare testing the homogeneity of the total variances was significant at the five per cent level for Spelling. The effect of race on the variances was negative i.e. Blacks were more variable than Whites. The races effect was also non-significantly negative for Arithmetic. There was no evidence for heterogeneity in the total variances. The chisquare testing the homogeneity of the total variances for Vocabulary approached significance and the races effect was positive (i.e. Whites were more variable than Blacks). The races effect was also positive for Heim, but the total variances were homogeneous.

The E_1E_2 model did not explain variation in performance on Verbal tests. Estimates of $\hat{E}_1\hat{G}_1$ and $\hat{B}$ obtained by fitting the $G_1B\,E_1$ model to Whites and Blacks were compared using a c-test. Estimates of $\hat{D}_R$ in Whites and Blacks obtained by fitting the E_1D_R model were also compared. These comparisons are summarised in Table A54.

The greater variability of Whites in Vocabulary cannot be attributed to any particular component of variation since estimates of $\hat{E}_1\hat{G}$ and $\hat{B}$ are all non-significantly greater in Whites. Estimates of genetical and environmental components do not differ significantly between Whites and Blacks for Heim or Arithmetic.

In the case of Spelling where Blacks were more variable than Whites, $\hat{E}_1$ and $\hat{B}$ are greater in Blacks. However, $\hat{E}_1$ is significantly greater in Blacks and appears to produce most of the heterogeneity in total variance. Factors producing differences in $\hat{E}_1$ between Blacks and Whites were discussed for Spatial tests.

TABLE A54: COMPARISON OF ESTIMATES OF GENETICAL AND ENVIRONMENTAL FACTORS BETWEEN RACES FOR VERBAL-ARITHMETIC TESTS

	Sign	S.L	E_1	G_1	B	D_R
Vocabulary	+	**	+0.193 NS	+1.086 NS	+0.606 NS	+2.214 *
Heim	+	NS	-1.498 NS	+0.821 NS	-0.071 NS	+0.621 NS
Arithmetic	-	NS	+0.561 NS	+0.477 NS	-1.851 NS	-1.403 NS
Spelling	-	*	-2.737 **	+0.576 NS	-1.676 NS	-1.398 NS

Note:

1. SIGN is the sign of the races effect
2. S.L is the significance level of the races effect
3. The figures in the body of the Table are values of c for comparing components between races.
4. If c is positive, the components are bigger in Whites than Blacks

   If c is negative, the components are bigger in Blacks than Whites

3.7 Discussion of the relative magnitudes of genetical and environmental components of variation in ability

Genetical theory makes no predictions about the proportion of the total variation that will be genetic in origin or the proportion that will be environmental. There is no theory of environmental causation which allows us to predict the relative magnitudes of different environmental components. However, we can predict, from genetical theory, the relative magnitudes of between and within families genetical components, given certain assumptions about the nature of the gene action.

Following Jinks and Fulker (1970), we redefine the between and within families genetical components of variation in terms of underlying gene effects as described in Section 3:

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

$$G_2 = D_R + \tfrac{1}{16} H_R + \tfrac{1}{2} \left[A/(1-A) \right] D_R$$

where D_R, H_R and A are defined as in Section 3.

We have data, only on twins reared together, so we can estimate $\hat{G}_1$, $\hat{E}_1$ and $\hat{B}$ where $\hat{B} = G_2 + E_2$. We may now specify the genetical parameters of our model in terms of gene effects:

$$\hat{G}_1 = \tfrac{1}{4} D_R + \tfrac{3}{16} H_R$$

$$\hat{B} = \tfrac{1}{4} D_R + \tfrac{1}{16} H_R + \tfrac{1}{2} \left(A/(1-A) \right) D_R + E_2$$

In the absence of assortative mating and E_2, we have a definitive test for dominance since we expect $G_1 > B$ by $\frac{1}{8} H_R$. Either assortative

mating or common environmental effects could make $\hat{B}$ significantly greater than $\hat{G}_1$ and we would be unable to distinguish between them unless evidence from another source were available.

The difference between $\hat{G}_1$ and $\hat{B}$ in these data was tested for significance for each of the fourteen tests, using a c-test. The results for Whites and Blacks separately are summarised in Table A55. Estimates of $\hat{B}$ are significantly greater than estimates of $\hat{G}_1$ for Verbal-Arithmetic tests except Vocabulary where the difference between $\hat{B}$ and $\hat{G}_1$ is not significant. For the Spatial tests we find that $\hat{B} > \hat{G}_1$. The difference between $\hat{B}$ and $\hat{G}_1$ is significant for Spatial ability, Formboard, Paper folding and Object aperture. In Blacks, $\hat{B} > \hat{G}_1$ for all tests except Identical pictures, Paper folding and Mazes completed which were anomalous in previous discussions. $\hat{B}$ is significantly greater than $\hat{G}_1$ for Spatial ability.

We have no evidence for dominance in these data since $\hat{G}_1$ is not significantly greater than $\hat{B}$ for any test. We cannot conclude that there is no dominance, since the contribution of any variance due to E_2 or assortative mating might increase the magnitude of $\hat{B}$ to a greater extent than the dominance increases the magnitude of $\hat{G}_1$. However, $\hat{B}$ is significantly greater than $\hat{G}_1$ for several tests. This demonstrates the effect of E_2 or assortative mating or both on variation among individuals in test performance. The effect of E_2 or assortative mating is also suggested for the other tests since $\hat{B}$ is greater than $\hat{G}_1$ in one race at least. In those cases where the difference between $\hat{B}$ and $\hat{G}_1$ is negative, the difference is not significant. Therefore, generally we may conclude that we have some evidence for assortative mating or common environmental effects for all measures of ability.

TABLE A55: COMPARISON BETWEEN $\hat{B}$ AND $\hat{G}_1$ FOR WHITES AND BLACKS

	VALUES OF C			
	WHITES		BLACKS	
Arithmetic	+2.197	*	+4.729	***
Vocabulary	-0.067	NS	+0.483	NS
Heim	+2.468	*	+2.264	*
Spelling	+1.982	*	+2.857	**
Calendar	+2.964	**	+2.649	**
Identical pictures	+3.730	***	-1.478	NS
Cubes	+1.710	NS	+0.462	NS
Surface development	+0.830	NS	+0.597	NS
Formboard	+2.852	**	+1.861	NS
Paper folding	+3.161	***	-1.406	NS
Object aperture	+2.528	*	+1.288	NS
Spatial ability	+3.051	**	+3.033	**
Mazes completed	+3.153	**	-0.311	NS
Mazes errors	+0.160	NS	+0.681	NS

NOTE: If c is positive, $B > G_1$

If c is negative, $G_1 > B$

If we are prepared to make further assumptions we may set limits to the possible values of E_2 and assortative mating. Assuming no dominance or assortative mating, we may obtain an estimate of common environmental variation:

$$\hat{E}_2 = \hat{B} - \hat{G}_1$$

The standard error is given by the square root of the variance which is found:

$$V(\hat{E}_2) = V(\hat{B}) + V(\hat{G}_1) - 2COV(\hat{B}\,\hat{G}_1)$$

Estimates of $\hat{E}_2$ and their standard errors are given in Table A56. These are underestimates of E_2 if there is dominance variation, but overestimates if there is assortative mating. The broad heritability may now be recalculated using these values of $\hat{E}_2$:

$$\hat{h}^2_b = \frac{2G_1}{2G_1 + E_1 + E_2}$$

Since $\hat{G}_1 = D_R + \frac{3}{16}H_R$, then in the presence of dominance, $\hat{h}^2_b$ is biased upwards by $\frac{1}{8}\ H_R/\frac{1}{2}D_R + \frac{3}{16}H_R + E_1 + E_2$. Therefore, this ratio may only be regarded as the heritability if there is no dominance.

The variance of $\hat{h}^2_b$ is given as the variance of a ratio as described in section 5.3.4. It turns out that the variance of $\hat{h}^2_b$ is:

$$V(h^2_b) = (h^2_b)^2\left[\frac{V_{G_1}}{G_1^2} + \frac{(4V_{G_1}+V_{E_1}+V_{E_2}+4W_{G_1E_2}+4W_{G_1E_1}+2W_{E_1E_2})}{(2G_1 + E_1 + E_2)^2} - \frac{(8\,V_{G_1}+4W_{G_1E_1}+4W_{G_1E_2})}{2G_1(2G_1 + E_1 + E_2)}\right]$$

Estimates of $\hat{h}^2_b$ for Whites and Blacks and their standard errors are given in Table A56.

Now, if we make the assumption of no dominance, but this time assume that there is no E_2, we can attempt to set a limit to the value of the assortative mating. In the absence of dominance and E_2, G_1 and B are defined:

$$G_1 = \tfrac{1}{4}D_R \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (1)$$

$$B = \tfrac{1}{4}D_R + \tfrac{1}{2}(a/(1-A))D_R \quad \ldots \quad \ldots \quad \ldots \quad (2)$$

Therefore, we have an estimate of $\hat{D}_R = 4G_1$. Substituting into equation 2, we may obtain an estimate of $\hat{A}$:

$$\hat{A} = \frac{B - G_1}{B + G_1}$$

We are assuming no dominance or E_2, therefore,

$$\hat{h}^2_B = \hat{h}^2_N = \frac{G_1 + B}{G_1 + B + E_1} = \frac{\tfrac{1}{2}D_R + \tfrac{1}{2}(A/(1-A))D_R}{\tfrac{1}{2}D_R + \tfrac{1}{2}(A/(1-A))D_R + E_1}$$

Fisher (1918) showed that A is a simple function of the marital correlation and the narrow heritability:

$$A = \mu h^2_N$$

Therefore, we can obtain estimates of the marital correlation:

$$\hat{\mu} = \frac{A}{h^2_N}$$

TABLE A56: ESTIMATES OF THE COMMON ENVIRONMENTAL COMPONENT AND THE BROAD HERITABILITY ASSUMING NO DOMINANCE OR ASSORTATIVE MATING

PART A

	WHITES				BLACKS			
	E_2	se	h^2_B	se	E_2	se	h^2_B	se
Arithmetic	152.409	84.821	0.552	0.100	404.162	136.533	0.349	0.113 *
Vocabulary	-0.105	2.041	0.553	0.147	0.898	2.442	0.279	0.273
Heim	81.543	31.914	0.566	0.095	113.667	61.013	0.384	0.020
Spelling	34.972	12.898	0.648	0.070	105.162	43.649	0.343	0.134 *
Calendar	15.547	6.653	0.082	0.184	20.923	3.452	-0.041	0.260
Identical pictures	89.097	28.857	0.325	0.102	104.657	74.340	0.126	0.256
Cubes	17.265	12.951	0.240	0.165	5.347	15.268	0.170	0.315
Formboard	15.723	6.956	0.167	0.151	3.151	4.658	-0.300	0.570
Paper folding	7.017	2.788	0.147	0.154	-4.677	4.412	0.841	0.260
Mazes completed	31.149	12.395	0.157	0.151	-5.737	20.610	0.532	0.231
Mazes errors	1.021	11.820	0.381		28.321	54.528	0.220	

PART B

TEST	GROUP	E_2	se	h^2_B	se
Surface Development	White males	18.445	5.751	0.113	0.256
	White females	-16.386	11.038	0.842	0.109
	Black males	48.735	16.933	-0.103	0.463
	Black females	-6.733	6.597	0.680	0.189
Object aperture	White males	17.074	5.835	0.194	0.230
	White females	5.997	2.740	0.218	0.218
	Black males	4.055	5.745	0.452	0.327
	Black females	3.257	1.833	0.099	0.354
Spatial ability	White males	169.041	43.268	0.283	0.141
	White females	64.554	30.339	0.599	0.098
	Black males	157.036	70.896	0.304	0.222
	Black females	81.367	27.195	0.397	0.136

Estimates of $\hat{A}$ and $\hat{\mu}$ will be overestimates if the assumption of no E_2 fails. However, failure of the assumption of no dominance will lead to underestimation of $\hat{A}$ and, therefore, of $\hat{\mu}$. Values of $\hat{G}_1$, $\hat{G}_2$, $\hat{E}_1$, $\hat{A}$, $\hat{h}^2_N$ and $\hat{\mu}$ are presented in Table A57. When we assume no dominance and no assortative mating and take $(\hat{B} - \hat{G}_1)$ as an estimate of $\hat{E}_2$ we find that the magnitude of $\hat{h}^2_B$ is considerably less than that of $\hat{Bh}^2_B$, the biased broad heritability. The proportion of the total variation due to genetical effects for Verbal-Arithmetic tests now lies between 0.55 and 0.65 in Whites and between 0.28 and 0.38 in Blacks. For Spatial tests the range of heritability estimates is greater, lying between 0.08 and 0.60 in Whites and 0.12 and 0.45 in Blacks. In several cases the estimates are meaningless since $\hat{G}_1 > \hat{B}$ (although not significantly greater) leading to a negative estimate of $\hat{E}_2$. The heritability of Blacks is, in most cases, non-significantly lower than that in Whites. This is true for all Verbal-Arithmetic tests and for most of the Spatial tests. Overall these estimates of heritability, obtained by assuming that common environmental effects alone contribute to $\hat{B} > \hat{G}_1$ are considerably lower than those reported in comparable studies.

Many workers have demonstrated a fairly strong correlation between the phenotypes of mates for measures of ability, lying between 0.3 and 0.5 (Burks, 1928; Willoughby, 1927, 1928; Spuhler, 1967). Thus, we expect that if there is a heritable component of variation that assortative mating may increase $\hat{B}$, as described in previous sections. Turning now to Table A56, we may assess values of $\hat{A}$ and $\hat{\mu}$ obtained by assuming no dominance and no E_2. Firstly, we note some anomalous results. For Vocabulary, Paper folding, Mazes completed and Surface development

TABLE A57: ESTIMATES OF GENETICAL AND ASSORTATIVE MATING PARAMETERS

TEST	GROUP	G_1	G_2	E_1	A	h^2_N	μ
Arithmetic	Whites	171.876	324.285	126.644	0.307	0.797	0.385
	Blacks	138.420	542.582	112.957	0.582	0.854	0.681
Vocabulary	Whites	3.133	3.028	5.174	-0.017	0.544	-0.031
	Blacks	1.151	2.049	5.055	0.281	0.388	0.724
Heim	Whites	36.195	167.738	50.856	0.321	0.833	0.385
	Blacks	57.521	171.188	71.055	0.497	0.763	0.651
Spelling	Whites	52.393	87.365	22.065	0.250	0.864	0.289
	Blacks	39.618	144.780	46.687	0.740	0.798	0.927
Calendar	Whites	1.670	17.217	21.770	0.823	0.465	1.770
	Blacks	-0.784	20.139	18.882	1.081	0.506	2.136
Identical pictures	Whites	35.783	124.880	59.373	0.555	0.730	0.760
	Blacks	18.351	123.008	150.150	0.740	0.485	1.526
Cubes	Whites	8.991	26.256	39.747	0.490	0.470	1.042
	Blacks	4.180	9.527	35.507	0.390	0.279	1.398
Formboard	Whites	3.675	19.398	20.907	0.681	0.525	1.297
	Blacks	-1.823	4.794	12.643	0.223	0.190	1.174
Paper folding	Whites	1.327	8.344	8.359	0.725	0.536	1.353
	Blacks	6.464	1.787	7.124	-0.567	0.537	-1.056
Mazes completed	Whites	6.394	37.543	36.656	0.709	0.545	1.301
	Blacks	21.400	15.663	43.397	-0.155	0.461	-0.199
Mazes errors	Whites	11.740	12.761	37.165	0.042	0.397	0.106
	Blacks	20.277	48.598	115.745	0.411	0.373	1.102
Surface development	White m	44.144	105.871	33.186	0.411	0.819	0.502
	White f	45.703	29.317	35.510	-0.218	0.691	-0.315
	Black m	-3.963	44.772	35.790	1.194	0.533	2.240
	Black f	16.080	9.347	21.864	-0.265	0.538	-0.493
Object aperture	White m	4.756	21.830	22.375	0.642	0.543	1.182
	White f	2.778	8.774	13.918	0.519	0.455	1.141
	Black m	5.421	2.345	6.345	0.254	0.606	0.419
	Black f	0.666	3.923	8.814	0.710	0.342	2.076
Spatial ability	White m	50.128	219.169	83.303	0.628	0.760	0.826
	White f	90.168	154.772	55.965	0.264	0.814	0.324
	Black m	46.491	203.526	56.232	0.628	0.816	0.770
	Black f	37.135	118.502	31.538	0.523	0.832	0.629

KEY: m = male

f = female

negative estimates of $\hat{A}$ and $\hat{\mu}$ have been obtained for one or more groups. $\hat{A}$ is negative because $\hat{B}$ is smaller than $\hat{G}_1$. For these tests $\hat{B}$ and $\hat{G}_1$ are not significantly different. The standard errors of $\hat{B}$ and $\hat{G}_1$ are large and $(\hat{B}-\hat{G}_1)$ might be negative by chance. A small proportion of anomalous results are expected in any study.

If we now consider the remaining tests several interesting observations can be made. A number of estimates of $\hat{\mu}$ for the Spatial tests are greater than one. This means that $\hat{A}$ is larger than $\hat{h}^2$. Since the marital correlation cannot be greater than 1.0, this implies that $\hat{A}$ is inflated by some other component. We notice that $\hat{\mu}$ is larger in Blacks than in Whites in most cases. Values of the marital correlation between spouses of about 0.5 have been observed in Whites (Vandenberg, 1972). The marital correlations in Blacks determined from the values of $\hat{A}$ calculated from these data are all much higher than this, suggesting that $\hat{\mu}$ and $\hat{A}$ are being inflated. In Whites the pictures is slightly different. Estimates of $\hat{\mu}$ for the Spatial tests are larger than expected, but for Verbal-Arithmetic tests, values of $\hat{\mu}$ are within the expected range (except for Vocabulary where $\hat{G}_1 > \hat{G}_2$).

Thus, we have three observations which we may attempt to explain: $\hat{A}$ is greater in Blacks than Whites for many tests; for Spatial tests $\hat{\mu}$ in both Blacks and Whites is greater than expected and for Verbal/Arithmetic tests $\hat{\mu}$ is larger than expected in Blacks, but slightly smaller than expected for Whites. Estimates of $\hat{A}$ are very sensitive to the value of $\hat{B}-\hat{G}_1$. We have already shown that where we have monozygotic and dizygotic twins reared together $\hat{B} = \frac{1}{4}D_R + \frac{1}{16}H_R + \frac{1}{2}(A/(1-A))D_R + E_2$. In estimating $\hat{A}$ and $\hat{\mu}$, we assumed no dominance or E_2. Dominance cannot inflate the size of $\hat{B}-\hat{G}_1$ since, as we have shown, $\hat{G}_1 = \frac{1}{4}D_R + \frac{3}{16}H_R$.

Therefore, inflation of $\hat{B-G_1}$, and hence $\hat{A}$ and $\hat{\mu}$, must be caused by E_2. Thus, we have some positive evidence for common environmental effects in Blacks and Whites on Spatial tests and in Blacks on Verbal-Arithmetic tests (but not for Whites on Verbal-Arithmetic tests), since $\hat{A}$ and $\hat{\mu}$ are greater than expected when we assume no E_2. However, although we can attribute the inflated estimates of $\hat{\mu}$ to the presence of E_2 in these data, we cannot make any firm conclusion about the source of the difference in $\hat{A}$ and $\hat{\mu}$ between Whites and Blacks. This difference may be a genuine difference in the degrees of assortative mating in the two races or it may be that the contribution of common environmental effects to variation is larger in Blacks. Although several studies have reported marital correlations for Whites, there are no reports of marital correlations for measures of ability in Blacks. Garrison et al (1968) report a study of over 100,000 marriages in Minnesota where the educational attainments of both parents were known. They summarise their findings in terms of the $\underline{G}$ statistic of Goodman and Kruskal (1954, 1963), which is a measure of association for data with non-Gaussian distributions. Values of $\hat{\underline{G}}$ for US Southern Whites and Blacks were 0.6640 ± 0.0008 and 0.6083 ± 0.0026. These values of $\hat{\underline{G}}$ are for educational attainments, but since a large correlation between educational attainments and other measures of ability has been found, we shall use these statistics for comparing Whites and Blacks. The value of $\hat{\underline{G}}$ for Whites is greater than that of Blacks, i.e. the degree of assortative mating is higher for Whites than Blacks for educational attainments. This lends no support to the hypothesis, suggested by the data of the Georgia Twin Study, that the degree of assortative mating for ability might be greater in Blacks. We note that the differences in variance

components and assortative mating parameters suggested by our data are small and non-significant and may be regarded only as tentative findings which need extensive re-testing with additional sets of data. Analysis of the data suggests that the unreliability of many of the individual tests is low. Therefore, on the basis of this one study, assuming that the degree of assortative mating is, if anything greater in Whites (from the study of Garrison et al) we may draw the tentative conclusion that the contribution of common environmental effect to variation in ability is greater in Blacks than in Whites.

3.8 Summary

In the analysis of the mean squares, we have been attempting to determine the most appropriate model for variation in the measures of ability being studied. We found that for Verbal-Arithmetic tests the most appropriate model was the genetical model since the simple environmental model failed. For the Spatial tests no firm conclusions could be drawn. The E_1E_2 model was adequate to explain variation for most tests. Thus, the data were consistent in most cases with the hypothesis that variation in Spatial ability can be attributed to environmental differences between individuals. When a simple two parameter genetical model was fitted to the data, this was also adequate to explain variation. Therefore, we could not discriminate between simple genetical and environmental models of variation, implying that the power of the tests were too low, and that a larger study would be needed in order to determine the source of variation between individuals in ability.

Martin (personal communication) has performed power calculations in an attempt to determine the size of experiment needed to discriminate between the E_1E_2 and E_1D_R models for the classical twin study, comprising monozygotic and dizygotic twins reared together. He assumed equal numbers of each type of twin and generated four statistics - the between and within pairs mean square for each type of twin - using arbitrary values for the parameters of the model. The number of pairs of twins needed to reject the simple genetical model at the 5% level in 95% of experiments when the simple environmental model was in fact the "correct" model was determined with a method involving the use of the non-central chisquare. Similarly the number of pairs of twins needed to reject the simple environmental model when the simple genetical model was the "correct" model was determined. When the data were generated assuming that $E_1 = E_2 = 0.5$ and the E_1D_R model was fitted to the data, Martin found that 430 pairs of twins (i.e. 215 pairs each of MZ and DZ twins) were required before the E_1D_R model failed. In the converse situation where the E_1E_2 model was fitted to data generated by assuming that $\frac{1}{2}D_R = E_1 = 0.5$, he found that 640 pairs of twins were required before the false model was rejected. In our study we have 364 pairs of twins. Therefore, even if the E_1E_2 model were the "true" model, we would still be unlikely to reject the E_1D_R model with a sample of this size. Conversely if the E_1D_R model were the "true" model and the heritability was 0.5 (i.e. $\frac{1}{2}D_R = E_1 = 0.5$), then we would not expect to be able to reject the simple environmental model. Martin also generated data where $\frac{1}{2}D_R = 0.8$ and $E_1 = 0.2$ and fitted the E_1E_2 model. In this case, only 118 pairs of twins were needed for the "false" E_1E_2 model to fail. Thus, as we would expect, the power of the test for a genetical component increases rapidly as the size of the genetical component increases.

Estimates of the biased broad heritability were about 0.8 for Verbal-Arithmetic tests and a significant genetical component was detected easily. The E_1E_2 model failed. The power calculations would have lead us to predict this for tests with a heritability of 0.8 and the sample size available here. The biased broad heritability for Spatial tests was somewhat lower, the average value being approximately 0.5. Thus, from the power calculations, we would have predicted that discrimination between E_1E_2 and E_1D_R models would be difficult. This is in fact what we found.

We have demonstrated a genetical component of variation for Verbal-Arithmetic tests, but have failed to discriminate between simple environmental and simple genetical models for Spatial tests. If we are now prepared to assume that both Verbal-Arithmetic and Spatial tests are in some measure under genetical control, but that we failed to detect a genetical component for Spatial tests because the power of the test was too low, we can then try to determine the source of difference in total variances between the two races. This does not seem to be an unreasonable assumption in view of the power calculations of Martin. Also, external evidence for a heritable component of variation in Spatial ability may be obtained from the studies of many workers (e.g. Block, 1968).

Comparing the estimated genetical and environmental components of variation in Blacks and Whites for Spatial tests, we saw that, in general, $\hat{E}_1$, $\hat{B}$ and $\hat{G}_1$ were all larger in Whites than Blacks, and therefore, the greater variability in Whites could not be explained by the greater magnitude of any particular genetical or environmental component of variation. The total variance of Whites was non-significantly less than that of Blacks for Identical pictures. Estimates of the genetical

components of variation were greater in Whites than Blacks. However, $\hat{E}_1$ was less in Whites than Blacks, the difference being highly significant. Therefore, the smaller variance in Whites can be accounted for by the smaller specific environmental component.

For Verbal-Arithmetic tests, the genetical component is larger in Whites than Blacks. The total variances for Spelling were significantly less in Whites. The $\hat{E}_1$ component was significantly less in Whites for this test.

When we attempted to estimate E_2/assortative mating, we found a difference between Verbal-Arithmetic and Spatial tests. $\hat{B}$ was bigger in Blacks for Verbal Arithmetic tests and smaller in Blacks for Spatial tests. The study of Garrison et al (1968) does not lead us to suppose that the degree of assortative mating is larger in Blacks than Whites, and, therefore, we tentatively propose that common environmental effects may be producing more variation in Blacks for Verbal-Arithmetic abilities. This could be the result of differences in educational opportunities between the two races.

Spatial tests have a high loading on 'g' and there is a lot of weight on Spatial tests in this battery. Thus, it is not surprising to find that $\hat{B}$ is larger in Whites than Blacks for the Spatial tests and the "general factor" of ability analysed in Section 7. Discussion of the significance of this finding will be postponed until Section 7, when it will be discussed in relation to the observed differences between the three sub-samples of the data, described in Section 1.5.3.

SECTION 6: GENOTYPE-ENVIRONMENT INTERACTION

1 INTRODUCTION

The simple empirical and theoretical models fitted to the mean squares in the last section all assumed additivity of genetical and environmental effects. Two types of non-additivity may be recognised:

(i) dominance and epistasis (i.e. genetical non-additivity)

(ii) genotype-environment interaction (G x E)

When simple models assuming no dominance were fitted to the data, we were testing the assumption of no dominance. Eaves (1972) has shown for several experimental designs, that the sample sizes needed for detecting dominance variation are large and many times the size of this study. Therefore, although theoretically dominance should produce failure of our simple additive models, in practise it is unlikely to do so with studies of this size.

We have shown that genotype-environment interactions are inevitably confounded with genetical and environmental components of variation with the present experimental design. Therefore, we do not expect the presence of G x E interaction to cause failure of our simple additive model for second degree statistics. However, the presence of genotype-environment interaction will bias estimates of genetical and environmental components of variation. If is remains undetected, we may seriously mislead ourselves about the magnitude and relative importance of the different components of variation. We shall therefore, use the scaling test proposed by Jinks and Fulker (1970) for the detection of systematic G x E. Both

systematic and unsystematic G x E contribute to variation i.e. to second degree statistics. However, only the former contributes to third degree statistics. Since the scaling test involves the regression of variances on means, only systematic G x E will be detected and unsystematic effects will remain undiscovered. Systematic genotype-environment interaction may be detected by means of our scaling test, but we will not be able to estimate the magnitude of the variation it produces, since our experimental design is not suitable for this purpose.

Genotype-environment interaction as a source of variation in human ability has been the subject of much controversy, especially since there are few studies of a suitable design for detecting it and estimating its magnitude. Several authors (e.g. Layzer, 1974; Lewontin, 1974) express the view that G x E interaction is probably an important factor in producing individual differences in ability, but that the problems involved in its detection and estimation are so profound as to preclude any worthwhile analysis of its effects in man. Many studies in plants and animals (e.g. Haldane, 1946; Mather and Jones, 1958; Bucio-Alanis, Perkins and Jinks, 1969; Jinks and Perkins, 1970; Jinks and Connolly, 1975; Mather and Caligari, 1975) have shown that genotype-environment interaction is, indeed, widespread and an important source of variation, although from their studies, we should not expect it to account for more than twenty per cent of the total variation. The methods applicable to its analysis in experimental animals and plants are not used in man, although the method for the detection of G x E given in Jinks and Fulker (1970) is based on one of the methods used for G x E analysis in experimental organisms. We shall follow the method of Jinks and Fulker for the detection of certain kinds of G x E interaction in the following analyses.

2 INDICATIONS OF GENOTYPE-ENVIRONMENT INTERACTION

In the previous section, the results of the model-fitting led us to postulate genotype-environment interactions, for some tests at least. G x E such that the size of the environmental component increases with the genotypic mean would lead to predictable consequences for the components of variation. The mean of Whites is greater than that of Blacks. Therefore, given interaction between the genotype and specific environmental variation, we would expect E_1 to be greater in Whites than in Blacks. This would in turn lead to a greater total variance in Whites. Thus, the presence of systematic genotype-environment interaction would explain both the greater total variation and the greater specific environmental component found in Whites.

The means and total variances were given separately by race, sex and zygosity, in Tables A17a and A17b for each of the fourteen measures of ability. One consequence of genotype-environment interaction of the type postulated above might be a correlation between the means and total variances of the eight groups of twins. We would expect those groups with higher means to have higher total variances given genotype-environment interaction. Such a correlation is often the first indication of systematic genotype-environment interaction. However, other factors could produce a correlation between means and variances. For example, if a trait is being measured more accurately at one end of the scale than the other, then the variance will be smallest where the trait is being measured most reliably and a correlation between means and variances will be produced.

Since we suspected genotype-environment interaction in these data, we looked at the means and variances and found a tendency for the means

and variances to increase together. We decided to perform a regression analysis to see if this effect was significant. The total variances of the eight groups of twins were regressed onto their means.

The regression mean square was significant for Vocabulary, Heim, Cubes, Surface Development, Formboard, Object Aperture and Spatial Ability. In all cases the total variance increased with the mean. These seven tests are the same seven whose total variances were shown to be heterogeneous by the Bartlett test (see Section 5), thus further suggesting some form of "genotype-environment interaction" for certain of the tests. Such a dependence of variance on mean could be produced by other types of genetical non-additivity, such as dominance and epistasis and we must proceed to a proper scaling test for G x E interaction.

However, first we will discuss the relationship between genotype-environment interaction and the problems involved in choosing a scale on which to measure our trait. Any scale used to measure a trait is in some sense arbitrary, and the choice of scale depends on our particular interest in the data. The "best" scale from the genetical point of view is the one which produces the simplest interpretation of the data. This means that we look for a scale on which all the variation may be explained by a simple additive model. In order to find such a scale we may have to transform our data. Any such transformation will change the parameter estimates of our model since different parts of the scale are altered to different extents and may also change the parameters needed to explain variation. For example we can choose a scale on which there is no need to invoke genotype-environment interaction, although there was genotype-environment interaction on the original scale.

Transformations removing undesirable non-additive influences are effectively changing the trait under study. If we wish to make predictions from our data, then we choose the scale with the simplest interpretation. However, it is possible that a particular scale is of intrinsic psychological interest to us. In this case we accept this scale and accept any complications that arise. Often the "problem" will be the presence of "genotype-environment" interactions, because on our psychologically meaningful scale we are not measuring the trait with equal precision in all parts of the scale. If there is some systematic component to the way in which our precision of measurement changes with scale, then it will be detectable as systematic genotype-environment interactions.

In these data we made no attempt to change the scale by transformation since we are interested in the analysis of the raw scores. We went ahead and fitted models to the mean squares based on the raw scores, i.e. we assumed no genotype-environment interaction. Normally the failure of this assumption would lead to failure of the model. However, when data on monozygotic and dizygotic twins reared together only is available, then G_1E_1 and BE_1 interactions are confounded with E_1, BE_2 with B and G_1E_2 with G_1. Therefore, estimates of the size of the interactions cannot be obtained, although we can ask how much of the environmental sum of squares is predictable from the genotype. However, the presence of interaction will not lead to the failure of models fitted to the mean squares.

Since the correlation between means and variances is highly suggestive of G x E interaction, we decided to investigate the possibility of interactions further. We may recognise two types of interaction:

a. systematic ("directional")

b. unsystematic ("ambidirectional")

Both types contribute to variation and, therefore, to second degree statistics. However, first and third degree statistics will only be affected by G x E if there is a directional component i.e. if the size of the environmental component changes in some systematic way with the genotype. Such directional G x E may be detected by the scaling test proposed by Jinks and Fulker (1970) for use in man, utilising third degree statistics, which they adapted from a similar test designed to detect systematic G x E in experimental organisms.

3 TEST FOR GENOTYPE-ENVIRONMENT INTERACTION

Jinks and Fulker (1970) have proposed tests for systematic GE_1 and GE_2 interactions using monozygotic twins. Monozygotic twins are identical genetically. However, one member of the pair will score higher than the other because of chance or environmental effects. The absolute difference between the scores of monozygotic twins may be used as a measure of the environmental effect within families. If the within families environmental influences (i.e. those contributing to E_1) are the same for all pairs of twins, then we expect the difference in scores between pairs of monozygotic twins reared together to be the same, apart from sampling error. If E_1 differs significantly from family to family, then either different families react differently to the same environmental influences or different families are subjected to environmental influences differing significantly between families. Pair sums are a measure of genotype and/or common family environment. Therefore, interaction between genotype (or E_2) and within family environment of a systematic kind, will lead to a regression of absolute pair differences on sum. Following a similar argument, a regression of absolute pair difference on sum for monozygotic twins reared apart will demonstrate systematic interaction between genotype and environmental differences between families (or within families).

Pair sums and absolute pair differences were calculated for all monozygotic twins, for each of the fourteen measures of ability. Theoretically, at least, the regression line of absolute pair differences on sum could take any shape. However, we decided to test only linear and quadratic relationships since higher order relationships would be

biologically and psychologically difficult to interpret. The analysis was performed using the polynomial regression analysis program (BMD05R) from the BMD package of programs (Dixon, 1973). This program tests the significance of any linear or quadratic component and has the facility for plotting the observed points and the predicted regression line.

The signs of the regression coefficients and their significance levels are shown in Table A57. We see that there is a significant linear regression for five tests, which (apart from the case of Identical Pictures) is accompanied by a significant quadratic component. Significant quadratic components were found for six tests altogether. For all tests, except Calendar, the linear component was positive and the quadratic negative. We have already commented on the high test specific variation found for the Calendar test, which led to the results for Calendar being inconsistent with those of the other tests in several of the analyses. The behaviour of the Calendar tests is anomalous once again since the shape of its regression line is completely different from the shapes for the other tests.

Regression coefficients are not easy to visualise or interpret. We, therefore, show the predicted regression lines for five tests in Figures A4 to A8, part a. These curves are all negative quadratics and the significant linear component found for some tests may be attributed to the asymmetry of the regression line. The quadratic component was significant for the tests depicted in Figures A5 to A8, but not for the Arithmetic test. The regression lines for all the tests except Calendar showed the same general shape, although the quadratic component was not significant for all tests. Thus genotype-environment interaction is suggested, such that extreme genotypes are less variable than intermediate genotypes.

TABLE A57: REGRESSIONS OF ABSOLUTE PAIR DIFFERENCES ON PAIR SUMS FOR ALL MONOZYGOTIC TWINS

	LINEAR		QUADRATIC	
	Sign	SL	Sign	SL
Arithmetic	+		-	
Vocabulary	+		-	
Heim	+		-	**
Spelling	+	**	-	***
Calendar	-		+	*
Identical Pictures	+	**	-	
Cubes	+		-	
Surface Development	+	*	-	**
Formboard	+	***	-	***
Paper Folding	+		-	†
Object Aperture	+			†
Spatial Ability	+		-	
Mazes Completed	+	**	-	***
Mazes errors	+		-	

Key:
† significant at the 10% level
* significant at the 5% level
** significant at the 1% level
*** significant at the 0.1% level
SL significance level
SIGN sign of the regression coefficient

FIGURE A4 THE ARITHMETIC TEST

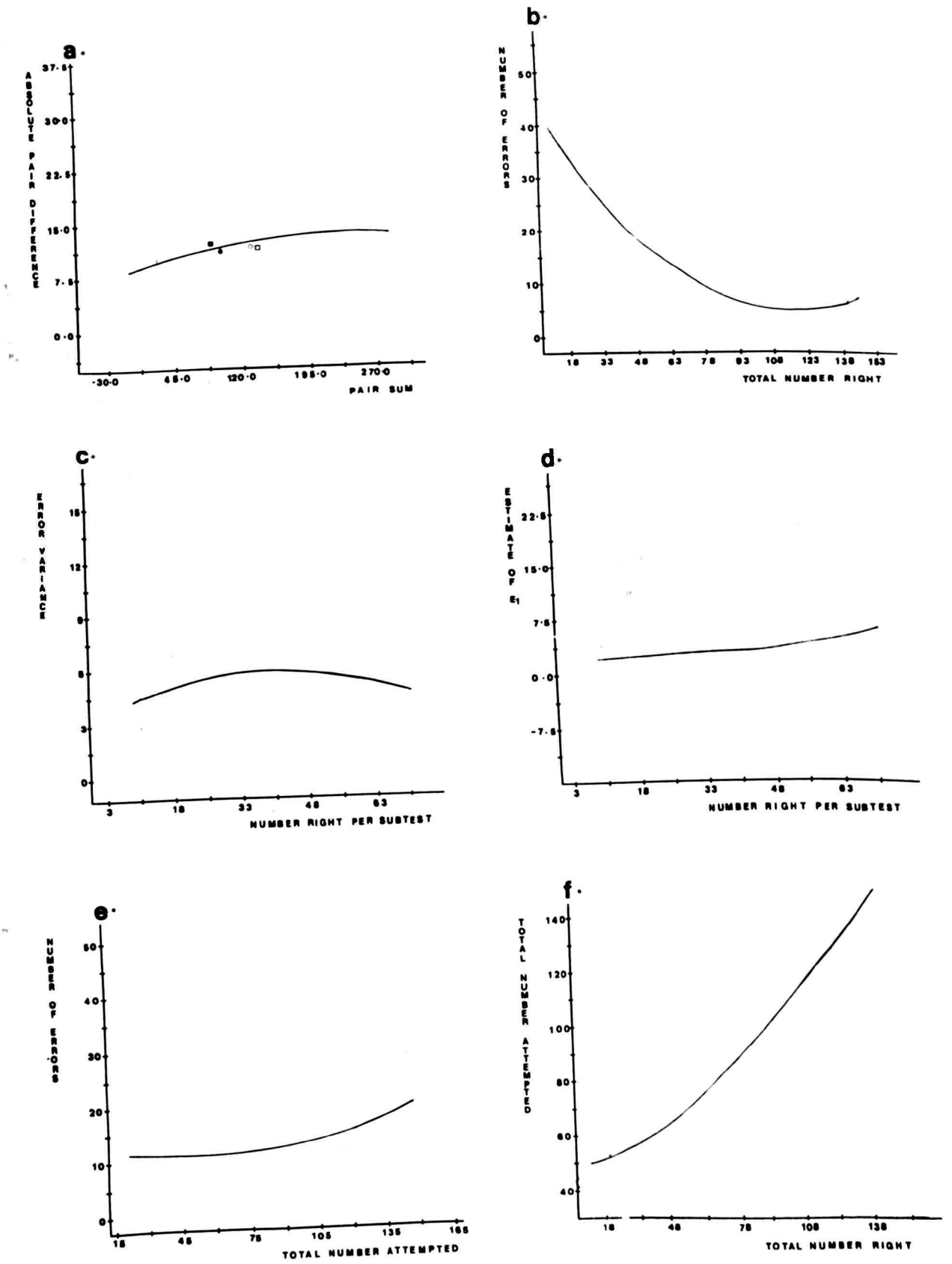
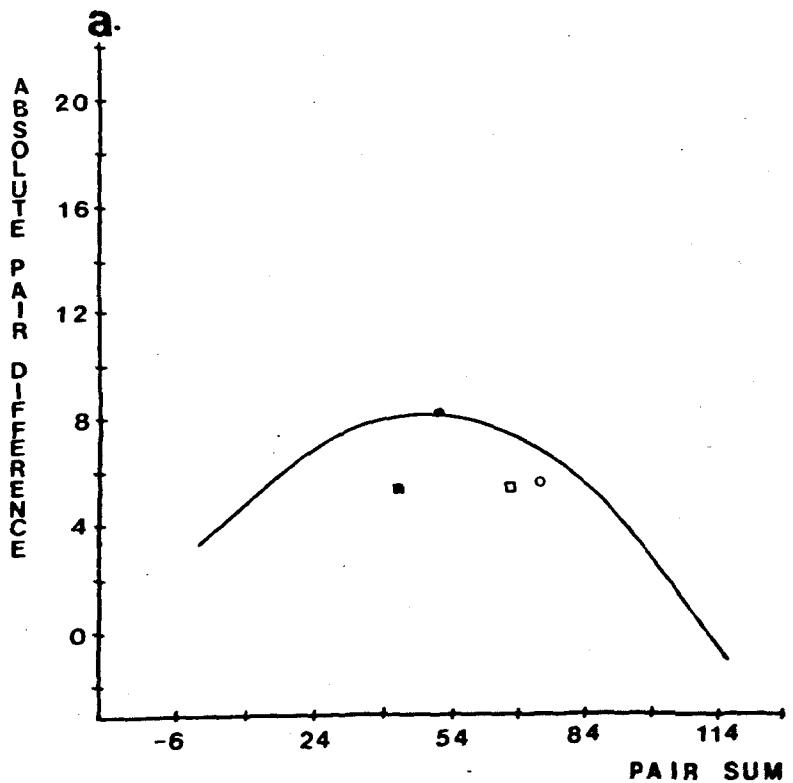


FIGURE A5

THE



SPELLING TEST

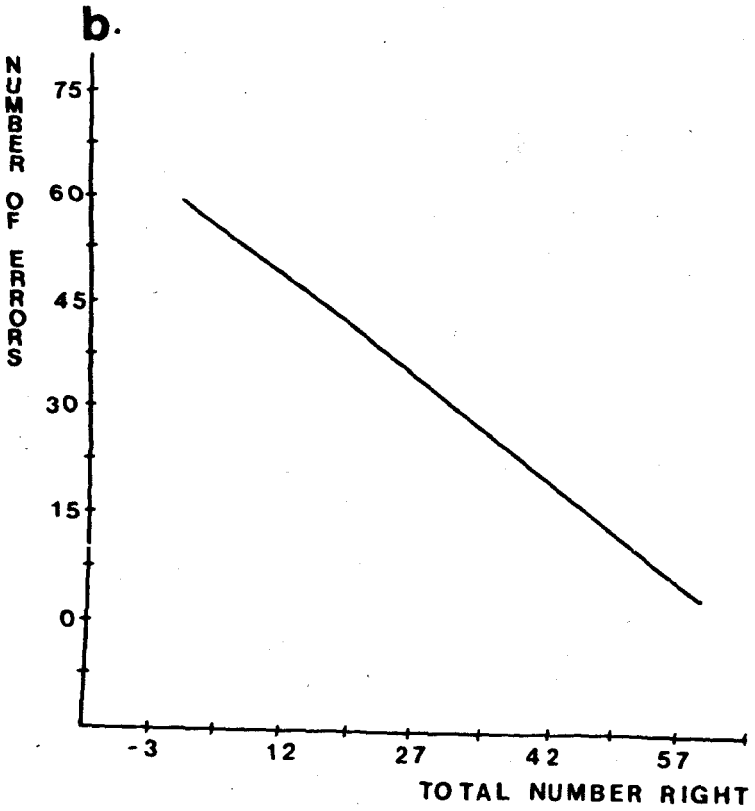


FIGURE A6 THE IDENTICAL PICTURES TEST

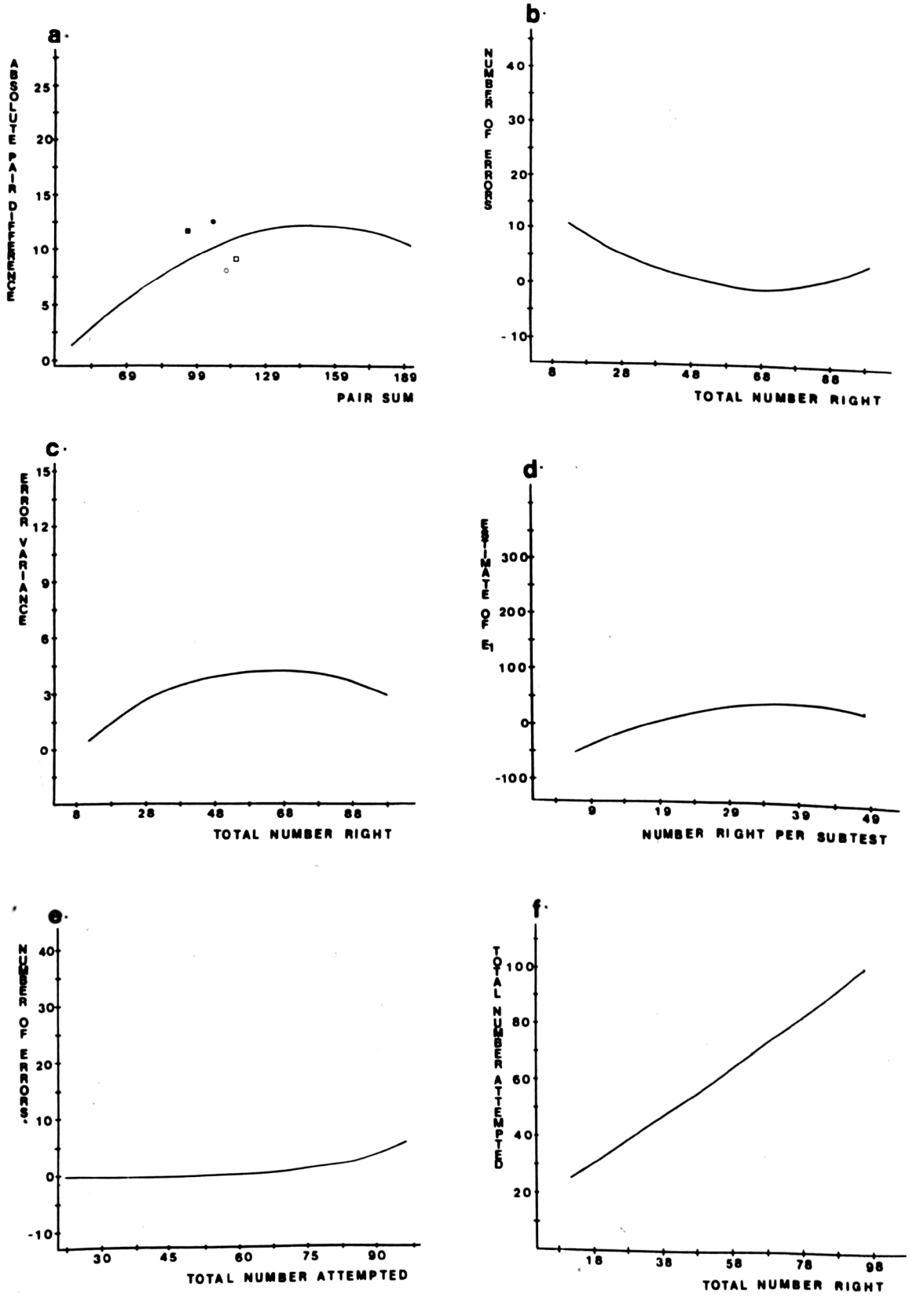


FIGURE A7 THE SURFACE DEVELOPMENT TEST

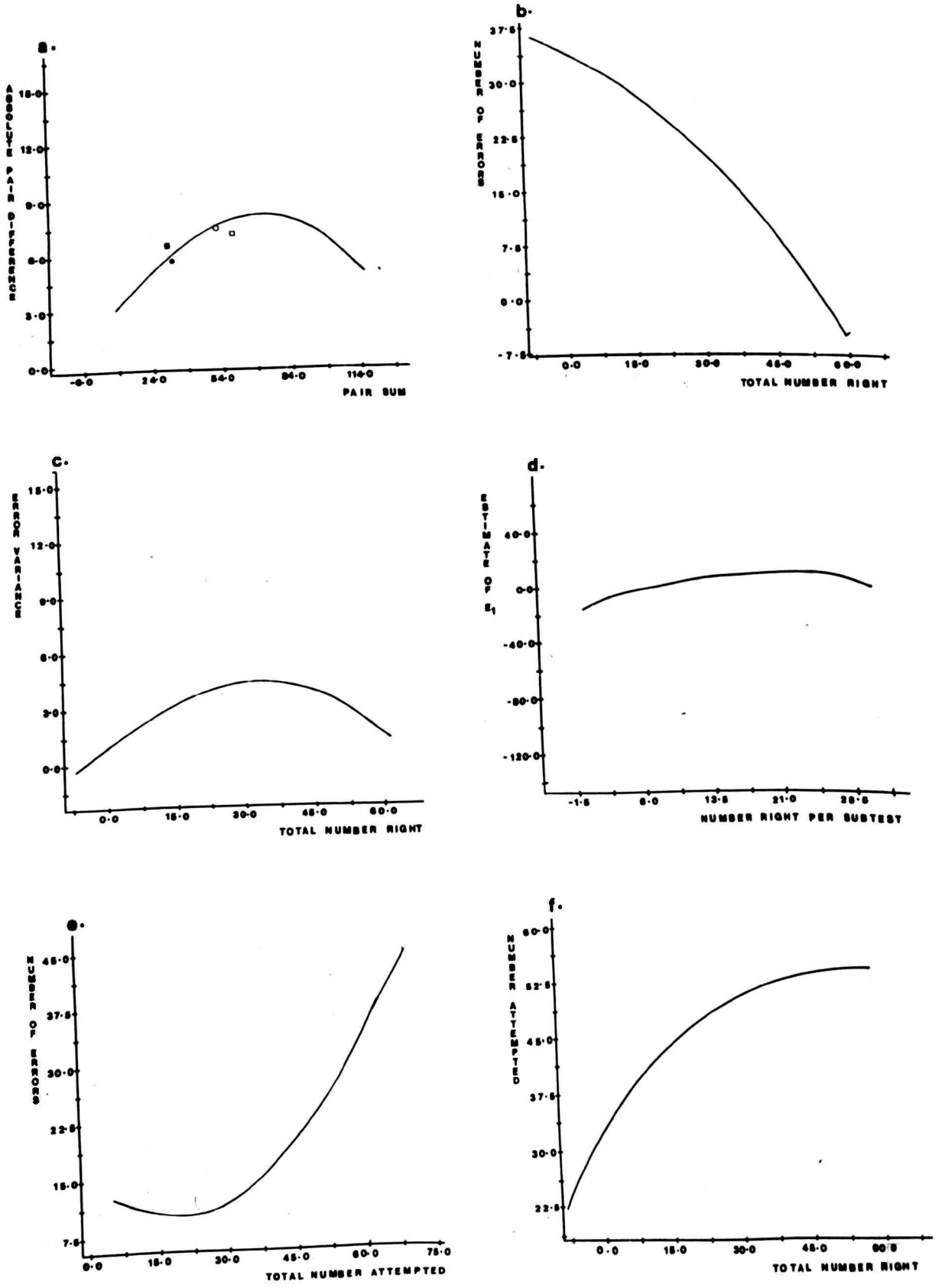
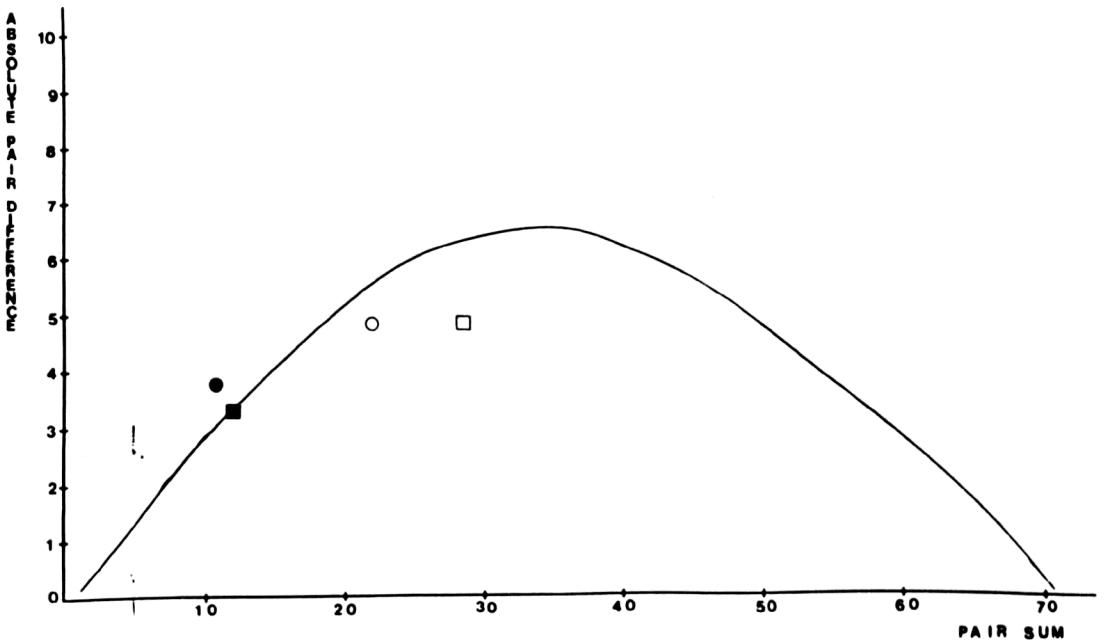


FIGURE A6.1 THE FORMBOARD TEST

REGRESSION OF INTRAPAIR DIFFERENCE ON PAIR SUM

a. FOR THE TOTAL SAMPLE



b. FOR EACH GROUP SEPARATELY

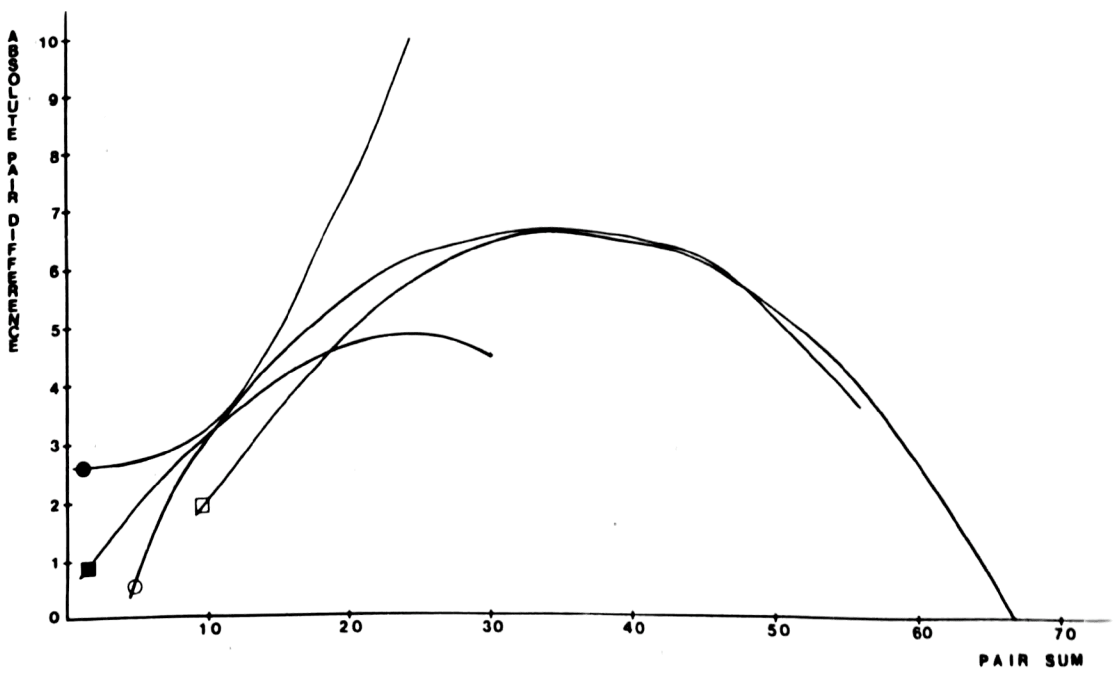
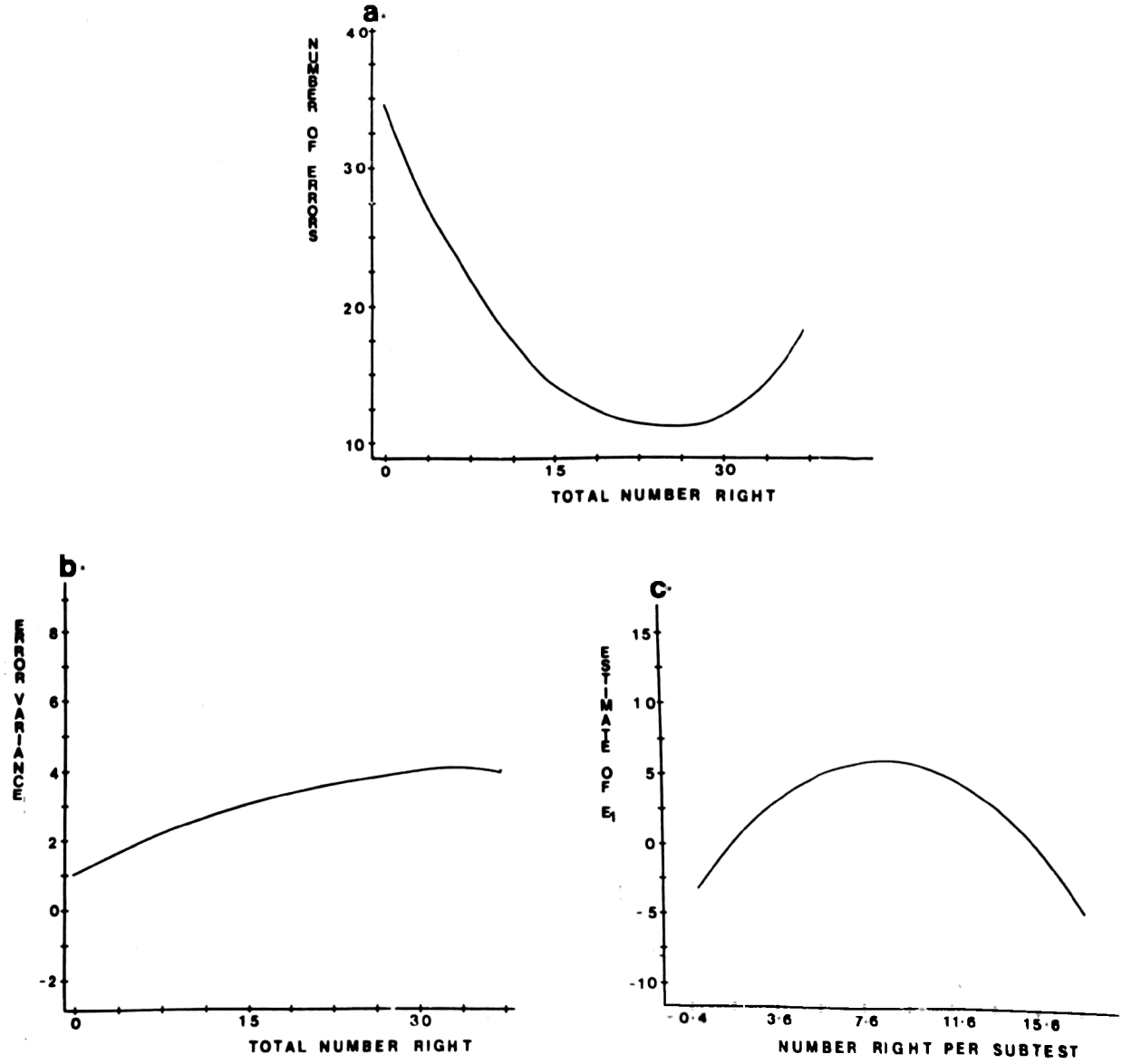
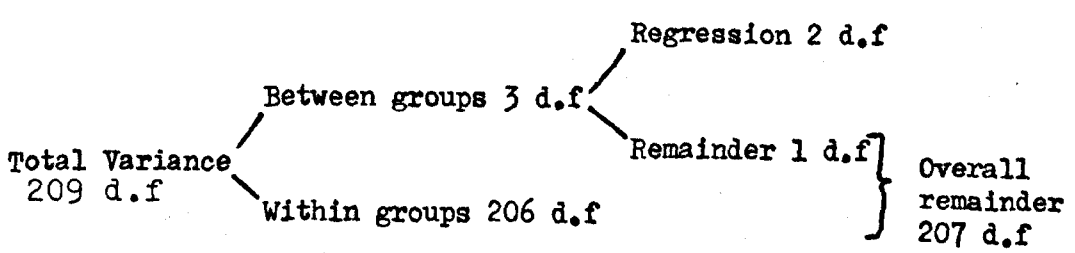


FIGURE A8.2 THE FORMBOARD TEST



We performed the same regression analysis separately by race and sex, since the means of the four groups differed and we had evidence from the model fitting that the within families environmental component differed from group to group. The signs of the regression components and their significance levels are shown in Table A58. The shapes of the regression lines and their regression coefficients differed between the four groups, with the regression coefficient even changing sign in some cases. This apparent inconsistency was hard to explain at first. However, when we plotted the curves for each of the four groups on the same axes, the reason for the inconsistencies became clear. Let us take Formboard as an example. The regression of absolute pair difference on pair sum for all monozygotic twins is shown in FigureA8.1a. The means of the four groups are indicated on the graph. In FigureA8.1b, the regression lines for each of the four groups are plotted separately on the same axes. We can now see that when the means differ, the differences in regression of absolute pair differences on sums may be attributed to the fact that different groups lie on different parts of the same curve. In order to confirm this subjective impression an analysis of variance between and within groups was carried out. We can break the data down in two useful ways. We can recognise the following structure in these data:



or alternatively:

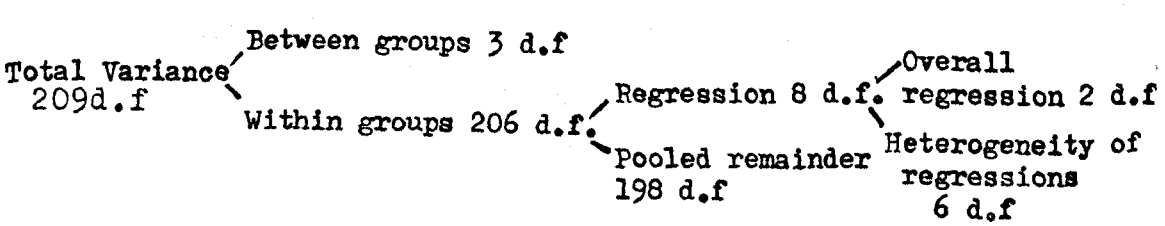


TABLE A58: REGRESSION OF ABSOLUTE PAIR DIFFERENCES ON PAIR SUMS BY RACE AND SEX

	WHITES								BLACKS							
	MALES				FEMALES				MALES				FEMALES			
	Linear		Quadratic		Linear		Quadratic		Linear		Quadratic		Linear		Quadratic	
Arithmetic	-	+	+		+		-		+		-	*	+		-	
Vocabulary	-		+		+		-		-		+		+		-	†
Heim	+	†	-		+		-	*	+		-		+		-	
Spelling	+	*	-	**	+	*	-	*	+		-	*	+		—	†
Calendar	-		+	*	-	*	+		-		+		-	*	+	
Identical Pictures	+		-		+		-		-	†	+		+	*	-	
Cubes	+		-		+	†	+		+		-		-		-	
Surface Development	+	†	-		+	*	-		-	**	+		-		+	
Formboard	+		-	**	+	**	-	*	+		-		-	†	+	***
Paper Folding	-		+		-		+		+		-		+	*	-	†
Object Aperture	+		-	†	+		+		+		-		+		-	
Spatial Ability	-	*	+		-		+		+		-		-		+	
Mazes completed	+		-		+		-	*	-	†	+		+	*	-	*
Mazes errors	+		-		+		-		-		+		+		-	

The overall heterogeneity sum of squares is a sum of squares for 9 degrees of freedom - 3 degrees of freedom from the between groups sum of squares and 6 from the heterogeneity of regressions sum of squares. This sum of squares was calculated by subtracting the pooled remainder sum of squares for 198 degrees of freedom from the overall remainder sum of squares for 207 degrees of freedom. The overall heterogeneity MS was tested for significance against the pooled remainder mean square. This analysis was repeated for each of the fourteen measures of ability. The overall heterogeneity mean square was not significant for twelve tests. However, for Surface Development and Identical Pictures it was significant at the five per cent level. Thus, generally, the differing regressions in the four groups of twins may be explained by the fact that the four groups lie on different parts of the same curve.

We can see that since the White means are higher than the Black means for these measures of ability and lie closer to the maximum point on the regression curve (see FigureA8.1a) that we might expect E_1 to be greater in Whites, given genotype-environment interaction of the type described above. We are postulating that any differences in environmental variation between groups may be explained as properties of the scale of measurement.

We discussed how genotype-environment interaction arises from the choice of a particular scale and we may now ask which properties of our scale are leading to the observed genotype-environment interaction. There are two possibilities:

1. The genotype-environment interaction is a property of the way in which the scale is constructed. For example, the trait is not measured with equal precision in all parts of the scale and so the reliability differs across the scale.

2. The genotype-environment interaction may be a fundamental property of the trait when measured on this scale. For example the tendency for subjects to guess may vary across the scale. Thus on this scale, the specific environmental component (E_1) may differ in different parts of the scale.

Both these possibilities are in some sense properties of the scale and may be removed by suitable transformation. It is likely that in the first case we will consider that our measuring instrument is poor and will change it either by changing the actual test structure or by transformation. However, in the second case we may decide that this scale is the most psychologically meaningful scale and try to estimate the variation produced by the genotype-environment interaction. Therefore, we shall now move onto a discussion of scales and examine the genotype-environment interaction in more detail.

4 PROPERTIES OF THE SCALE

For several of our tests there is a regression of within pairs variability upon genotypic mean, such that extremes are less variable. We now wish to determine whether this relationship is produced by the different reliability of measurement in different parts of the scale, or whether there are genuine differences in sensitivity to the environment between individuals of different genotypes when measured on this scale. Therefore, we shall examine some factors producing unreliability of measurement to see if these can account for our observed regression.

Various models of the behaviour of an individual when faced with a test item have been formulated. Two scales are commonly used to measure ability. These are: (in the notation of Lord and Novick, 1968)

(i) number right = x_a

(ii) "formula score" = $x_a - \frac{w_a}{A - 1}$

where x_a = the number right

w_a = number wrong

A = number of categories

The first model assumes that an individual's knowledge may be measured by the number of correct responses he makes. No account is taken of correct responses obtained by guessing or whether wrong responses were obtained by incorrectly answering due to wrong information or omitting items. Clearly this is a very simple and unrealistic model, although it often used for scoring measures of ability.

The second model which gives rise to the formula score is known as the random guessing model. We assume that if a subject knows the answer to a test item, he gives that answer and if he does not know the answer he omits the item or guesses at random. Clearly this model is also simple and unrealistic. Usually a subject will have either misinformation or else partial information about an item which will render some particular response more attractive to him, especially in multiple-choice type tests. However, since all the fourteen tests studied here are either "number right" or "formula scores", we shall consider how the simple random guessing model described above gives rise to the formula score.

If we assume the random guessing model, a correction for guessing can be made to the number right score. The problem is to try and determine the distribution of scores if no random guessing had occurred i.e. we are trying to find a subjects true level of knowledge (k_a) from his observed number of items attempted (n_a) by making a correction for the number of items for which the correct answer was obtained by guessing. Lord and Novick (1968) show how the formula score, $\hat{k}_a$, is an approximation to the maximum-likelihood estimator of k_a. The formula score is given by $\hat{k}_a = x_a - \frac{w_a}{A - 1}$

where $w_a = n_a - x_a$

The formula score, $\hat{k}_a$, is an unbiased estimator of k_a (when n_a is fixed) over repeated independent random guessing. When there are no omitted responses i.e. n_a is the same for all subjects, then the formula scores is completely correlated over subjects with the number right score.

The error variance of a formula score can be divided into two-additive parts; the topastic error variance - due to guessing and the scedastic error variance - error variance due to all other causes. The topastic error variance of a number right score is:

$$\text{Var}\,(x_a \mid n_a, k_a, p) = (n_a - k_a)\,pq$$

where p = probability of obtaining a correct response

and q = 1-p

Thus, following Lord and Novick (1968), the variance of the formula score is:

$$\text{Var}\,(k_a \mid n_a, k_a) = (n_a - k_a)\,\frac{p}{q} = \frac{n_a - k_a}{A - 1}$$

i.e. for both number right scores and for formula scores the error variance is proportional to the number of errors, since $w_a = n_a - k_a$. Thus, we expect the error variance to increase with the number attempted (n_a) and to decrease with the level of knowledge (k_a). What consequences does this have in terms of our tests of ability and the observed genotype-environment interaction for many of the tests? In order to answer this question, we shall set up a model of test behaviour based on the error variance of the formula scores given by Lord and Novick.

5 A POSTULATED MODEL OF TEST BEHAVIOUR

We have seen how the topastic error variance of a score is proportional to the number of errors (w_a). Since $w_a = n_a - k_a$, we must attempt to explain our observed genotype by within families variation interaction in terms of the number of items attempted and the level of knowledge of the subjects. As measured ability (i.e. x_a) increases, we postulate that n_a increases more rapidly than k_a so that as ability increases, the number of items attempted increases even faster until a ceiling is reached. For tests showing genotype-environment interaction of the type found here, the difference between number attempted and number right should be largest in the middle range of ability, and the maximum size of this difference may vary from test to test. Thus we expect low scores to reflect not merely wrong answers, but also few attempted questions. The greater error variability will apply to subjects of intermediate ability who, we predict, will attempt many questions but know relatively few answers. As we progress to subjects of high ability, we expect to find most items attempted and answered correctly. Consequently, we expect a reduction in error variance as we move from subjects of intermediate to those of high ability. This model suggests that test scores may be explained in terms of two underlying traits; knowledge (k_a) and tendency to attempt questions (n_a). We may regard these two underlying parts of test behaviour as reflecting ability and motivation respectively. If the model is correct, then it is possible that the underlying traits may contribute differently to final performance in the different groups of twins (i.e. between races and/or sexes) and that we may be looking at different traits in the different

groups. If this is the case, then the heterogeneity of the total variances found for some tests should not surprise us. In this case analysis of the two underlying traits might be more useful in explaining differences between the groups. However, before speculating about the consequences of looking at different traits in the different groups, we should attempt to test our model.

Our hypothesis is that the model of test behaviour assumed by the number right scores and formula scores is the correct one and that the error variance is directly proportional to the number of errors. We predict that the gehotype-environment interaction will be explained by the error variance varying in different parts of the scale such that the error variance is proportional to the number of errors that the number of errors will vary in a similar manner in different parts of the scale. In order to test these predictions we shall look at the regressions of error variance and numbers of errors on final score.

6 TESTING THE POSTULATED MODEL OF TEST BEHAVIOUR

6.1 The Additional Information required for Testing the Model

The scoring systems for the fourteen measures of ability used in the Georgia Twin Project are shown in Table A59.

In order to test the predictions made in the previous section, additional information on each test was required. The within pairs variance was used as a measure of error variance plus genuine E_1 when regressing absolute pair differences on pair sums. We now wish to determine whether the regression of within pairs variance on pair sums was produced by differing error variances in different parts of the scale or whether it was produced by the interaction of genotype with within pairs environmental variance. Therefore, we need a measure of test unreliability.

The number of sub-tests comprising each test is shown in Table A59. We can obtain a measure of reliability for each of the eight tests with two or more sub-tests if we have the scores of individuals on the sub-tests. We do not have this information for the Mazes tests, the three Verbal tests or the Calendar test so we shall not be able to determine the source of the regression of absolute pair differences on pair sums for those tests. In order to test our prediction that the model for number right and formula scores was correct and that the error variance was proportional to the number of errors, we needed more detailed information about test performance. We needed to know the number attempted and the number right. Dr. R.T.Osborne kindly made available to us, in the Spring of 1975, the more detailed information needed to pursue the analysis:

TABLE A59: SCORING FORMULAE AND AVAILABILITY OF INFORMATION FOR TESTING OUR POSTULATED MODEL

Test	Scoring Formula	x_a and n_a availability	Number of Sub-Tests
Arithmetic	$x_a - (wa/4)$	✓	7
Vocabulary	$x_a - (wa/4)$	✓	1
Heim	$x_a - (wa/5)$	✓	1
Spelling	x_a	✓	1
Calendar	$x_a - w_a$	✓	1
Identical Pictures	$x_a - (wa/4)$	✓	2
Cubes	$x_a - w_a$	✓	2
Surface Development	x_a	✓	2
Formboard	x_a	✓	2
Paper folding	$x_a - (wa/4)$	✓	2
Object aperture	$x_a - (wa/4)$	✓	2
Spatial ability	x_a	✓	6
Mazes completed	x_a	×	1
Mazes errors	x_a	×	1

Note: For Mazes errors x_a = number of errors

(i) the number of items attempted by each individual on each test - n_a

(ii) the number of items correctly answered by each individual on each test - x_a

(iii) the score of each individual on each sub-test of the eight tests with two or more sub-tests

The information about n_a and x_a were not available for the Mazes tests and so the analysis of the genotype-environment interaction cannot be taken any further for these tests.

6.2 The Regression of Error Variance and E_1 on Test Score

It is possible, for those tests with two or more sub-tests, to determine separately the effects of within families variation due to error variance (test unreliability) and that due to genuine E_1 differences between individuals. We shall use the analysis for a test with two sub-tests as an illustration. We wish to obtain an estimate of the error variance and E_1 at each point on our scale represented by a pair of twins. Therefore, the analysis of variance outlined in Table A60A was performed separately for each pair of twins for the six tests with two sub-tests. The same analysis, with the appropriate numbers of degrees of freedom, was also performed for Spatial Ability and Arithmetic which have six and seven sub-tests respectively. In the analysis the between twins item is a measure of the variation within pairs produced by both true error variance and by E_1 differences between individuals. The model for the expected mean squares has been written assuming that both main effects are

TABLE A60: ANALYSIS OF VARIANCE FOR A PAIR OF TWINS

A. THE ANALYSIS WE PERFORMED ASSUMING THAT SUBTESTS ARE RANDOM

Item	df	Expected Mean Square
Between Twins (T)	1	$\sigma^2 + \sigma_I^2 + 2\sigma^2_T$
Between Sub-Tests (s)	1	$\sigma^2 + \sigma_I^2 + 2\sigma^2_s$
T x S	1	$\sigma^2 + \sigma_I^2$

B. THE ANALYSIS NEEDED IF SUBTESTS SHOULD BE A FIXED EFFECT

Item	df	Expected Mean Square
Between Twins (T)	1	$\sigma^2 +$ $+ 2\sigma^2_T$
Between Sub-Tests (s)	1	$\sigma^2 + \sigma_I^2$ $+ 2\sigma^2_s$
T x S	1	$\sigma^2 + \sigma_I^2$

random. Since we have no replicates, we have to assumr that the σ_I^2 interaction component is not significant and use the twins x sub-tests mean square as our estimate of error. Then we can find σ_T^2, the variance within a pair of twins corrected for unreliability:

$$\sigma_T^2 = \frac{T - (T \times S)}{2}$$

where the divisor of two is the number of sub-tests. The component σ_T^2 is a measure of the differences between individuals after the error variance has been removed and is, therefore, our estimate of E_1 for a pair of twins. The error term, T x S is our estimate of the error variance for a pair of twins.

There are two problems with this analysis:

(1) There is no test for the interaction component since we have no replicates. Therefore, if there is a singificant interaction component, our estimate of the error variance will be an over-estimate. This will not bias E_1 as long as both main effects may be regarded as random.

(2) It is difficult to decide whether sub-tests are a fixed or a random effect. If we were to test the twins again to obtain replicates we would probably use the same sub-tests on a later occasion. This suggests that sub-tests are a fixed effect. However, we would do this for convenience only. It is probably more realistic to regard sub-tests as random samples from a universe of possible sub-tests. Therefore, it is performance on all possible sub-tests that we are really interested in and we may regard sub-tests as a random effect.

If sub-tests were fixed, the model in Table A60B would be appropriate and if there was a significant interaction component, then our estimate of E_1 would be an underestimate.

Having obtained estimates of error variance and E_1 for all pairs of twins on each of the eight tests, we tried to determine which of these effects is the source of the regression of absolute pair differences on pair sums. A polynomial regression analysis of the error variances on the number right was performed using the BMD program. Some examples of the predicted regression lines are shown in Figures A4c, A6c, A7c and A8.2b. The polynomial regression of our estimates of E_1 on number right per sub-test was also performed. Examples of the predicted regression lines are shown in Figures A4d, A6d, A7c and A8.2c The sizes and signs of the regression coefficients were examined and the shapes of the curves were assessed by looking at graphs of the predicted regressions. Our assessment of how well the regressions of error variance and E_1 mimic the regression of absolute pair differences on pair sums is summarised in Table A61. For six of the tests the regression of error variance on test score was more similar to the regression of absolute pair difference on sum than the regression of E_1 on test score, indicating that the regression of absolute pair difference on sum is being produced by different error variances in different parts of the scale rather than by differing sensitivities to the environment of genotypes in different parts of the scale. However, for Formboard and Spatial Ability, the regression of absolute pair difference on sum was best explained by differing sensitivities to the environment in different parts of the scale.

To summarise our findings in this Section, we see that for most tests the "genotype-environment interaction" may be explained as a property of the scale of measurement, since the regression of error variance on test score seemed the most likely explanation of the regression of absolute pair difference on pair sum. For Formboard and Spatial ability,

TABLE A61: COMPARISON BETWEEN REGRESSION LINES OF ABSOLUTE PAIR DIFFERENCE, ERROR VARIANCE AND $\hat{E}_1$ ON SOME MEASURE OF GENOTYPE

	Regression on measure of genotype of:	
	Error Variance	$\hat{E}_1$
Arithmetic	1	0
Identical Pictures	2	1
Cubes	2	0
Surface Development	2	1
Formboard	1	2
Paper Folding	2	1
Object Aperture	2	0
Spatial Ability	0	2

Key

0	No relationship between g x e and the shape of this regression line
1	The g x e and this regression line have the same sign on the quadratic component
2	This regression line follows the shape of the g x e very closely

there was some evidence that the "genotype-environment interaction" could not be explained by differing error variances in different parts of the scale and might be the result of genuine differences in within families environmental effects in different parts of the scale.

We shall now discuss the second part of our model which predicts that the model of test behaviour assumed by the number right and formula scores is correct and that therefore, the error variance is directly proportional to the number of errors.

6.3 The Appropriateness of the Number Right and Formula Scores

The topastic error variance is directly proportional to $n_a - k_a$ (i.e. the number of errors), if the models of test behaviour underlying the number right and formula scores are appropriate. Given that these models are appropriate, we predict that the regression of w_a on x_a should be the same shape as the regression of absolute pair (difference)^2 on pair sum for the six tests where the "genotype-environment" interaction could be explained as a property of the scale of measurement. This regression should be a quadratic:

$$w_a = a + b\,x_a + c\,x_a^2 \quad \text{where a, b and c are constants}$$

Alternatively it follows, since $w_a = n_a - x_a$, that:

$$n_a = a + (1 + b)\,x_a + cx_a^2$$

In order to test these predictions, we performed a polynomial regression analysis of w_a on x_a and n_a on x_a. Examples of these regressions are shown in Figures A4b, A4f, A5b, A6b, A6f, A7b, A7f and

A8.2a. We also thought it would be of interest to regress w_a onto n_a, since the algebraic relationship between w_a and n_a is complex and if the model of test behaviour is correct we may wish to study the two underlying traits "ability" and "motivation", discussed in Section 6.5. Examples of these regressions are given in A4e, A6e and A7e.

The regression of w_a onto x_a was the expected quadratic for only Cubes and Object Aperture. It followed quite closely the regression of absolute pair difference on pair sum for these two tests. However, for all the other tests the number of errors decreased as the number right increased. For seven tests the rate of change of w_a with x_a decreased as x_a increased and for five tests the rate of change increased as x_a increased, but apart from the case of Cubes and Object Aperture, the quadratic component was not significant and the linear component was highly significant. We do not see the expected increase in number of errors as ability increases from low to intermediate levels, and then a decline in number of errors as ability increases from intermediate to high levels. We see only that number of errors decreases as ability increases. The regression of error variance on x_a is a negative quadratic of the same shape as the regression of absolute pair differences on pair sum. The regression of w_a on x_a is not a quadratic but has a marked negative linear component. Therefore, our model of test behaviour inherent in the number right and formula scores, predicting that the error variance is directly proportional to the number of errors, cannot be accurate for most tests. In order to find further evidence on this point, we shall now examine the regressions of n_a on x_a, which would be negative quadratics on our model.

The regression of n_a on x_a was a negative quadratic for Cubes and Object Aperture. For all other tests, the linear component was highly significant and n_a increased as x_a increased. In the case of the regression of w_a on n_a there was a strong linear component such that w_a decreased as n_a increased. Thus from these regression lines, we do not find any evidence that predictions based on the properties of the number right and formula scores have any generality for these tests.

The regression of absolute pair differences on pair sums is best explained by differing environmental effects in different parts of the scale for Formboard and Spatial Ability. Thus we have some evidence for genuine genotype-environment interaction for two tests. The biases introduced into the estimates of genetical and environmental components of variation obtained in Section 6, were extensively discussed in Section 3. An additional point must be made here. For those tests where the regression of absolute pair differences on pair sums was not explained by genuine "genotype-environment interaction", but by differing error variances in different parts of the scale, then genetical and environmental parameters of variation will still be biased by the interaction generated by the properties the scale. These biases will be the same as those discussed in Section 3, since estimates of E_1 always contain error variance as well as true within families environmental variance.

7 SUMMARY

Tests where the regression of absolute pair differences on pair sums cannot be explained by "g x e" are best explained by differing error variances in different parts of the scale. For two tests i.e. Cubes and Object Aperture, the random guessing model specified by the formula scores is an adequate explanation of test behaviour, as far as the evidence presented here can show. However, for the other tests, predictions based on the formula scores are clearly not borne out by the data. The distribution of the number of errors is not the same as the distribution of total error variance. The random guessing model is not an adequate description of test behaviour for all tests i.e. subjects display misinformation, which cannot be attributed to guessing. Examination of Appendix II shows that, on a purely descriptive level, the tests are very different in their construction and the type of subject response required and we should not be surprised that one psychometric model is not adequate to explain test behaviour for all these tests, especially so simple a model as the random guessing model. Thus our general conclusion is that the test battery is a collection of heterogeneous tests for which different models of test behaviour and different models for individual differences in performance are required (see Section 5). However, we have seen how grouping the tests into two factors, one Spatial and one Verbal has helped in drawing fairly general conclusions about the data. In the next Section, we shall extract a measure of "general ability" from the various tests of the battery and attempt to draw more general conclusions about the causes of variation than has been possible through analysing individual tests.

SECTION 7: ANALYSIS OF A "GENERAL FACTOR"

1 INTRODUCTION

Several general conclusions and trends have emerged from analysing each measure of ability separately. Interpretation was greatly facilitated by recognising the division of the the tests into Verbal - Arithmetic and Spatial Factors. Tests of the two factors behaved differently in several respects. Tests of the Verbal-Arithmetic factor gave quite consistent results, although the answers obtained for Vocabulary did not always agree with those for the other tests. The amount of test specific variance for Vocabulary was greater than for the other tests (see Table A12). This might partly explain the divergence of the results for Vocabulary from those for the other tests. Although there was an overall similarity in the results obtained for the different Spatial tests, there were many inconsistencies between them. From Table A12, we can see that there is a high level of test specific variation for many of these tests especially for the Mazes tests and Identical Pictures. This might produce inconsistencies in the results obtained over different tests.

In order to obtain a more clear cut and simple interpretation of the data than was possible by analysing each test separately, we decided to try and extract a general measure of overall "ability". Spearman's (1904) theory of ability postulates a general factor, underlying all intellectual tasks, and specific factors, s, which produce variation in specific intellectual abilities. The fourteen tests of the Georgia Twin Study all purported to measure some aspect of ability, but the high level

of test specific variation made interpretation of the data difficult. Extraction of a general factor of ability should allow us to make more useful statements about the causes of individual differences in ability in Whites and Blacks and in males and females.

2 EXTRACTION OF THE GENERAL FACTOR

Principal components analysis was used to extract a general factor of ability. Principal components analysis is briefly described in Section 2.1 and Appendix C and a more extensive account may be found in Seal (1964). The first principal component extracted from a body of data accounts for the largest proportion of the variation and may be regarded as an approximation to a general factor accounting for variation common to all tests. In a large battery of tests, this first component is analogous to Spearman's g.

The principal components analysis was carried out using the raw age-adjusted scores. Standardised scores were not used since standardisation removes differences in total variances. Such differences in total variance are often the first indication of interesting differences between different groups of data, e.g. a difference in variance between males and females often indicates sex-linkage. One of our chief interests in analysing the general factor scores is to throw more light on the differences in variance components between Whites and Blacks.

We will analyse general factor scores from a principal components analysis based on all the data jointly. This is important since if we wish to compare the two races, we must assume that the same psychological model is appropriate for both races. In performing the principal components analysis separately for the two races, we would be assuming that a different model was possible in the two races. Analysing factor scores based on a principal components analysis carried out separately for each race would be equivalent to assuming that a different psychological model was required in each race, and, therefore, when we made comparisons

between races we would essentially be comparing the genetical and environmental components of variation for different traits.

All calculations necessary to complete the principal components analysis were carried out by Dr. R.T.Osborne, Professor of Psychology at the University of Georgia on their IBM 360/370 computers. Factor scores were assigned to each subject on the basis of the first principal component extracted from the age-corrected raw scores of the combined White and Black groups. These factor scores (FT's) were made available to me for further analysis by Dr. Osborne during 1975.

Osborne also completed a similar principal components analysis for White and Black groups separately, so that some test of whether the same model is appropriate for both races could be made. He found the factor loadings based on the first principal component from the analysis of Whites separately (which we shall call FW's). He repeated this procedure for Blacks and we shall call these factor loadings the FB's. Then he found the correlation between the FT's of Whites and their FW's and the correlation between the FT's of Blacks and their FB's. This correlation was 0.999749 in the case of Whites and 0.997475 in the case of Blacks. These extremely high correlations between the factor loadings based on the total group and the factor loadings based on each race separately suggest that the same psychological model is appropriate for both races. Thus, the factor scores made available by Dr. Osborne are an approximation to a general factor of ability which we believe to be the same general factor in both races. So we can analyse these factor scores in the same way as we analysed the raw, age-corrected test scores and make comparisons between sub-groupings of the data.

3 PRELIMINARY ANALYSIS OF THE GENERAL FACTOR SCORES

3.1 The Means

The means of the general factor scores were calculated separately by race, sex and zygosity. They are shown in the first column of Table A62. The simple linear model predicting the effects of race, sex and zygosity on the group means was fitted to the data as described in Section 2. Estimates, obtained from fitting this model, and their standard errors are presented in Table A63. The differences in mean between the eight groups of twins may be explained in terms of a significant races and zygosities effect. The races effect is positive i.e. the mean score of the Whites is greater than that of the Blacks. Races effects for the individual tests were also significant and positive. Whites score higher than Blacks in both "general ability" as measured by general factor scores, and in tests of specific abilities. Possible causes of the higher White mean in ability were discussed in some detail in Section 2.5.2.

We found for the individual tests that males scored higher on average on tests of Spatial ability than females and discussed the possibility in Section 3, that variation in Spatial ability might be produced by sex-linked genes. The strongest evidence we found for sex-linkage was the higher variance of males compared with females but several other explanations were consistent with the data.

Female means were higher than male means for the Arithmetic and Spelling tests, but not for Vocabulary or Heim. From the individual tests we could not predict the effect of sex on the general factor scores.

TABLE A62: SUMMARY STATISTICS FOR THE GENERAL FACTOR SCORES

Group			Mean	Total Variance	Between pairs mean square
WHITES	MALES	MZ	199.17436	4634.1077	8847.559
		DZ	191.42513	4477.9189	7727.945
	FEMALES	MZ	198.24893	3041.5923	5696.941
		DZ	186.35255	3341.0293	5558.672
BLACKS	MALES	MZ	138.66743	3289.9708	6461.645
		DZ	126.97304	2477.2525	3187.786
	FEMALES	MZ	144.2511	2737.5518	5220.191
		DZ	135.10973	2594.1306	4363.586

Within pairs mean square	F	Signi-ficance levels	Number of pairs
470.815	18.792	***	84
1291.624	5.983	***	51
416.188	13.688	***	87
1150.431	4.832	***	82
240.281	26.892	***	26
1817.471	1.754	0.15	14
304.563	17.140	***	50
878.294	4.968	***	33

TABLE A63: ESTIMATES OF EFFECTS OBTAINED BY FITTING A LINEAR MODEL TO THE MEANS FOR THE GENERAL FACTOR SCORES

	Estimate	se	c	significance level
Mean	165.039	1.024	161.171	***
Race	28.781	1.024	28.106	***
Sex	-0.939	1.024	0.917	NS
Zygosity	5.039	1.024	4.921	***
RxS	0.244	1.024	0.238	NS
RxZ	-0.131	1.024	0.128	NS
SxZ	-0.186	1.024	0.182	NS
RxSxZ	-0.853	1.024	0.833	NS

From Table A63 we see no evidence for a mean difference between males and females in general ability as measured by the general factor scores.

The effect of zygosity on the means of the general factor scores is highly significant. The mean of monozygotic twins is higher than that of dizygotic twins. This repeats a similar finding for many of the individual tests. In Section 2.5.4 we discussed evidence from Bulmer (1970) showing that the dizygotic twinning rate is higher in mothers from lower socio-economic status groups and we showed how this could account for the lower mean scores of dizygotic twins.

3.2 The Variances

The total variances for each of the eight groups of twins are shown in the second column of Table A62. The total variances of several of the individual tests were heterogeneous. Whites tended to be more variable than Blacks and for the Spatial tests; males tended to be more variable than females. The chisquare from Bartlett's test, testing the assumption that the total variances are all estimates of the same σ_T^2, was significant at the one per cent level ($\chi_7^2 = 19.21$), i.e. the total variances were heterogeneous.

In order to determine the source of this heterogeneity, the approach of Section 5.2.1 was adopted. The simple linear model predicting the effect of race, sex and zygosity on the total variances was fitted to the logarithms of the total variances using the weights specified in Section 5.2.1. The estimates of the effects and their standard errors are shown in Table A64. The heterogeneity in the total variances may be explained by a difference in total variance between the races. Whites

TABLE A64: ESTIMATES OF EFFECTS OBTAINED BY FITTING A LINEAR MODEL TO THE LOG VARIANCES OF THE GENERAL FACTOR SCORES

	Estimate	se	c	Significance level
Mean	8.084	0.059	136.460	***
Race	0.162	0.059	2.728	**
Sex	0.106	0.059	1.797	NS
Zygosity	0.035	0.059	0.586	NS
RxS	0.072	0.059	1.216	NS
RxZ	-0.050	0.059	0.838	NS
SxZ	0.045	0.059	0.756	NS
RxSxZ	-0.013	0.059	0.214	NS

are more variable than Blacks. A difference in total variance in this direction was found for many of the individuals tests. Heterogeneity of total variances between the sexes was found for several Spatial tests, which was our only evidence for sex-linkage in these data. The value of c testing whether the sex effect is significantly greater than zero for the log variances of the general factor scores approached significance ($c = 1.797$, $p < 0.10$). Thus, although the sex effect is not fully significant, there is some indication that the total variances of males are higher than those of females.

In fitting models to all the data jountly, the implicit assumption is made that the total variances are homogeneous. Factors contributing to heterogeneity in the total variances were described in detail in Section 5. When one of these assumptions fails and the total variances are heterogeneous, models will fail when fitted to all the data jointly. The assumption that the total variances are equal has failed in these data. Therefore, since we may attribute this heterogeneity to heterogeneity between races, we will fit models to Whites and Blacks separately. However, we shall fit models to the sexes jointly since the sex effect was not significant.

3.3 Genotype-Environment Interaction in the General Factor

The assumption of homogeneity of the total variances has failed for the general factor scores. This problem will be overcome by fitting models to races separately. The models specified in Section 3 assume that there is no interaction between genotypes and environment. We will test this assumption now, using the test suggested by Jinks and Fulker (1970) (see Section 6).

Absolute pair differences in general factor score were regressed onto pair sums using the polynomial regression analysis program from the BMD package (Dixon, 1973). We tested for linear and quadratic relationships since the biological meaning of higher order relationships would be difficult to interpret. The regression was performed for all monozygotic twins jointly and then separately by race and sex. There was no significant linear or quadratic relationship between absolute pair differences and pair sums, either for all monozygotic twins jointly or in any race-sex group. Thus, we have no reason to suspect the presence of genotype-environment interaction for the general factor, although there was "G x E" for several of the individual tests. Each individual test is measuring a slightly different aspect of ability i.e. a different trait with different scalar properties. We were not surprised to find interaction for some traits but not others. The general factor measures a trait of general ability common to performance on all the individual tests and does not explain test specific variation. Since we have found no evidence for genotype-environment interaction for this trait, we will fit the additive models specified in Section 3 to the mean squares of the general factor scores. We notein passing that there is no test for interaction between genotype and the between families environmental component. The biases that would be introduced by the presence of G x E_2 interaction were discussed in detail in Section 6.

3.4 Analysis of Variance

The total variances were partitioned into between and within pairs mean squares by a simple one-way analysis of variance, performed separately

by race, sex and zygosity. These mean squares are presented in columns four and five of Table A62. Values of F and the corresponding significance levels are also given. Mean differences between pairs were highly significant in all groups except dizygotic black males. The total sample size in the case of dizygotic Black males was only 14 pairs and, therefore, we were not surprised that differences between pairs were not significant, since the power of the test was low.

4 MODEL FITTING TO THE MEAN SQUARES

4.1 The Simple Environmental Models

The expectations for the E_1 and E_1E_2 models were given in Table A27. The E_1 model failed in both races (χ_7^2 = 216.778 in Whites, χ_7^2 = 84.971 in Blacks). This means either that there is significant between families variation or that there are significant differences between the mean squares of monozygotic and dizygotic twins.

We next fitted the E_1E_2 model. This model was not adequate to explain variation (χ_6^2 = 43.192 in Whites, χ_6^2 = 42.903 in Blacks). The E_1E_2 model specifies that variation may be explained in soley terms of environmental differences between families.

4.2 The Empirical Genetical Model

The G_1, B, E_1 model was fitted to the mean squares. Estimates of the parameters and their standard errors are given in Table A65. This model was adequate to explain variation in both races. The biased broad heritabilities ($\hat{Bh}_b^2$) and their standard error are also given in Table A65.

The estimate of biased broad heritability, $\hat{Bh}_b^2$, of the general factor is larger in Blacks than in Whites. However, the difference between the two heritabilities is not significant (c = 0.612). Therefore, we may regard the heritabilities as estimates of a value of Bh_b^2 common to Whites and Blacks.

Estimates of $\hat{E}_1$ and $\hat{B}$ are significantly larger in Whites than Blacks. (c = 2.437 for $\hat{E}_1$ and c = 2.119 for $\hat{B}$). However, the estimate of $\hat{G}_1$ is

TABLE A65: THE $G_1 B E_1$ MODEL FOR THE GENERAL FACTOR SCORES FOR WHITES AND BLACKS SEPARATELY

	WHITES				BLACKS			
	Estimate	Se	c		Estimate	Se	c	
G_1	763.380	153.990	4.957	***	902.607	243.861	3.701	
B	2614.128	300.265	8.706	***	1609.188	367.224	4.382	
E_1	442.849	47.850	9.255	***	281.655	45.663	6.168	
χ^2_5	6.351				4.647			
Bh_b^2	0.884	0.015			0.899	0.020		
E_2	1850.748	383.891	4.821		706.581	532.319	1.327	
h_b^2	0.3996	0.0642			0.6462	0.1390		
A	0.548	0.093			0.281	0.199		
μ	0.620	0.106			0.313	0.222		

non-significantly smaller in Whites than Blacks (c = 0.483). Since estimates of genetical and environmental components of variation differ in the two races, the fact that the biased heritabilities are not significantly different must be due to chance. This shows that "heritability" is not a very useful statistic. Since the two estimates of $\hat{Bh}_b^2$ do not differ, we might suppose that there are no racial differences in components of variation between races which are of interest to us. These differences are not apparent from estimates of "heritability" in the two races.

The total variances of Whites are greater than those of Blacks, and this may be attributed to the greater E_1 and B in Whites. The greater E_1 may indicate that the general factor is being measured less reliably in Whites, or it may mean that there is a greater range of specific environmental variation in Whites. If the reliability of measurement of the general factor is less in Whites, we might expect the test reliabilities of the individual tests to be less in Whites. Since we have no data on the individual test reliabilities in Whites and Blacks we cannot discriminate between these alternatives.

Our estimate of $\hat{B}$ is larger in Whites. Since $B = \hat{G}_2 + E_2$, either G_2 or E_2 or both may be larger in Whites. If we assume no dominance or assortative mating, then $G_1 = G_2$. Since $\hat{G}_1$ is less in Whites than Blacks and $G_1 = G_2$, then $\hat{B}$ must be larger in Whites because E_2 is larger, given no dominance or assortative mating. If there is assortative mating, then $\hat{B}$ may be larger in Whites because the level of assortative mating is higher. We cannot say why $\hat{B}$ is larger in Whites until we have tried to assess the relative importance of common environmental effects and assortative mating.

4.3 Common Environmental Effects and Assortative Mating

G_1 and B may be defined in terms of the underlying genetical effects:

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

$$B = \tfrac{1}{4} D_R + \tfrac{1}{16} H_R + \tfrac{1}{2} (A/(1-A)) D_R + E_2$$

Therefore, in the absence of dominance, assortative mating and common environmental effects, we expect $G_1 = G_2$. From Table A65, we see that $\hat{B}$ is greater than $\hat{G}_1$ in both races. The difference between $\hat{B}$ and $\hat{G}_1$ is significant in Whites ($c = 5.485$) but fails to reach significance in Blacks ($c = 1.603$). If there is dominance but no E_2 or assortative mating, then we expect $G_1 > B$ by $\frac{1}{8} H_R$. Since $\hat{G}_1$ is not greater than $\hat{B}$ for the general factor, we have no evidence for dominance. This does not mean that there is no dominance. The presence of E_2 or assortative mating inflates G_2 compared with G_1. If E_2 and assortative mating contribute to $\hat{B}$ to a greater extent than dominance contributes to $\hat{G}_1$, i.e. if $\frac{1}{8} H_R < \frac{1}{2} (A/(1-A))D_R + E_2$, then $\hat{G}_1$ may not be less than $\hat{B}$, even when there is dominance.

Setting Limits to the Size of the Common Environmental Effect

We expect assortative mating and common environmental effects to inflate B. Since $\hat{B} > \hat{G}_1$, we have evidence for either assortative mating or E_2 or both. If we make several simplifying assumptions, we may set limits to the values that E_2 and the assortative mating parameter may take.

Assuming no dominance or assortative mating, we can estimate E_2:

$$\hat{E}_2 = \hat{B} - \hat{G}_1$$

$$= (\tfrac{1}{4}D_R + E_2) - \tfrac{1}{4} D_R$$

Values of $\hat{E}_2$ were found as the difference between $\hat{B}$ and $\hat{G}_1$. The variance of $\hat{E}_2$ is given by:

$$V(\hat{E}_2) = V(\hat{B}) + V(\hat{G}_1) - 2W(\hat{B}\,\hat{G}_1)$$

The values of $\hat{E}_2$ for Whites and Blacks and their standard errors are given at the bottom of Table A65. If the assumption of no assortative mating fails, then $\hat{B} - \hat{G}_1 = \frac{1}{2} A/(1 - A)\, D_R + E_2$ and our estimate of $\hat{E}_2$ is an overestimate. If the assumption of no dominance fails, $\hat{B} - \hat{G}_1 = E_2 - \frac{1}{8} H_R$ and our estimates of $\hat{E}_2$ is an underestimate.

The estimate of $\hat{E}_2$ is significant in Whites, but not in Blacks. The value is non-significantly greater in Whites than Blacks ($c = 1.743$, $p < 0.10$).

The broad heritability may be calculated:

$$\hat{h}_B^2 = \frac{2\hat{G}_1}{2\hat{G}_1 + \hat{E}_1 + \hat{E}_2}$$

This estimate is unbiased if there is no dominance. Otherwise, it is biased by $\frac{1}{8} H_R/(\frac{1}{2}D_R + \frac{1}{4} H_R + E_1 + E_2)$. The broad heritabilities for Whites and Blacks and their standard errors are shown beneath the estimates of E_2 in Table A65. The heritability in Blacks is non-significantly larger than in Whites ($c = 1.61$, $p < 0.10$). From Table A65, we see that this heritability estimate is larger in Blacks because $\hat{G}_1$ is larger and $\hat{E}_2$ smaller in Blacks.

Setting Limits to the Size of the Assortative Mating Parameter

Let us now attempt to set a limit to the value that the assortative mating parameter, A, may take. Assuming no dominance or common environmental effects, we may write $\hat{G}_1$ and $\hat{B}$:

$$\hat{G}_1 = D_R \qquad \ldots(1)$$

$$\hat{B} = D_R + \tfrac{1}{2}(A/(1-A))D_R \qquad \ldots(2)$$

Therefore, we may estimate $\hat{D}_R = 4\hat{G}_1$. Substituting into equation (2), we find that:

$$\hat{A} = \frac{\hat{B} - \hat{G}_1}{\hat{B} + \hat{G}_1}$$

Following the approach outlined in Section 5.3.7, and assuming no dominance or assortative mating:

$$\hat{h}_B^2 = \hat{h}_N^2 = \frac{G_1 + B}{G_1 + B + E_1} = \frac{\frac{1}{2} D_R + \frac{1}{2} (A/(1-A))D_R}{\frac{1}{2} D_R + (A/(1-A))D_R + E_1}$$

and the marital correlation, μ, may be estimated as:

$$\mu = \frac{A}{h_n^2}$$

If the assumption of no E_2 fails, $\hat{A}$ and $\hat{\mu}$ will be overestimates and the value of the heritability will be an overestimate. Failure of the assumption of no dominance will lead to the underestimation of $\hat{A}$ and $\hat{\mu}$. The estimate of heritability will no longer be the narrow heritability, since the variance due to dominance will appear in the numerator. However, it is still an estimate of the broad heritability, given that there is no E_2.

Estimates of $\hat{A}$ and $\hat{\mu}$ are shown for Whites and Blacks at the bottom of Table A65. The variance of the ratio $\hat{B}-\hat{G}/\hat{B} + \hat{G}_1$ is given by:

$$V(\hat{A}) = \frac{V(\hat{B}-\hat{G}_1)}{\hat{B} + \hat{G}_1} = \frac{4}{(\hat{G}_1 + \hat{G}_2)^4} \left(\hat{B}^2 V(\hat{G}_1) + \hat{G}_1^2 V(\hat{B}) - 2\hat{G}_1 \hat{B} W \hat{G}_1 \hat{B}\right)$$

and the variance of the ratio $\frac{\hat{A}}{\hat{h}^2}$ is given by:

$$V(\hat{\mu}) = V\left(\frac{\hat{A}}{\hat{h}^2}\right) = \hat{\mu}^2 \left(\frac{V(\hat{A})}{\hat{A}^2} + \frac{V(\hat{h}^2)}{(\hat{h}^2)^2}\right.$$

The standard errors of $\hat{A}$ and $\hat{\mu}$ given in Table A65. Values of $\hat{A}$ and $\hat{\mu}$ are non-significantly larger in Whites than Blacks ($c = 1.216$ for $\hat{A}$ and $c = 1.248$ for $\hat{\mu}$).

Discussion of Common Environmental Effects and Assortative Mating

We have shown that $\hat{E}_1$ is significantly greater in Whites, but that the smaller $\hat{G}_1$ in Whites is not significantly different from that in Blacks. Either E_2 or the level of assortative mating is greater in Whites than Blacks. We have no expectation for the absolute magnitude of E_2 and, therefore, we cannot say whether the estimates of E_2 are reasonable or not. However, our values of $\hat{h}_b^2 = \frac{2G_1}{2G_1+E_1+E_2}$ (which are biased upwards by $\frac{\frac{1}{8}H_R}{2G_1+E_1+E_2}$ if there is dominance). ($h_b^2 = 0.4$ in Whites and $h_b^2 = 0.6$ in Blacks) are much smaller than those found for general ability in other studies.

This gives us some indication that our estimates of $\hat{E}_2$ may be inflated by variation produced by assortative mating, especially in Whites.

We will now reverse this argument and consider the size of $\hat{A}$ and $\hat{\mu}$. Our estimates of $\hat{A}$ and $\hat{\mu}$ are derived from the estimates of genetical parameters. However, other workers have measured the marital correlation directly and derived estimates of $\hat{A}$. The study of educational attainments of Garrison et al (1968) did not lead us to expect large differences between the marital correlations of Whites and Blacks. Several other studies of general ability (e.g. Vandenberg, 1972) lead us to suppose that our value of $\hat{A}$ obtained for Whites is rather high, but that the value of $\hat{A}$ of 0.281 obtained for Blacks is about right. This suggests that the value of $\hat{A}$ may be inflated by E_2 in Whites but that E_2 is negligible in Blacks.

Let us now assume that the true value of A is the same in both Whites and Blacks and that the true value is that of Blacks i.e. $\hat{A} = 0.281$. In Section 3 we showed that:

$$\hat{A} = \frac{G_2 - G_1}{G_2 + G_1}$$

and, therefore, we may estimate G_2:

$$\hat{G}_2 = \frac{G_1(1 + A)}{(1 - A)} = 1360.0692$$

in Whites, and therefore, $\hat{E}_2 = 1254.059$.

The estimates were now rewritten in Table A66. These estimates were calculated, assuming no dominance and that A(WHITES) = A(BLACKS) = 0.281. Given that these assumptions are reasonable, we see that the genetical parameters are approximately equal in the two races, but that $\hat{E}_1$ is significantly larger and $\hat{E}_2$ non-significantly larger in Whites than Blacks. Since the environmental parameters are larger in Whites, the

TABLE A66: PARAMETER ESTIMATES ASSUMING THAT ASSORTATIVE MATING IS THE SAME IN WHITES AND BLACKS

	WHITES	BLACKS
G_1	763.380	902.607
G_2	1360.069	1609.188
E_1	442.849	281.655
E_2	1254.059	-
h^2_B	0.556	0.899
A	0.281	0.281
μ	0.505	0.313

heritability of Whites is smaller than that of Blacks. This is interesting since several other workers have found the reverse (e.g. Scarr-Salapatek, 1971). The lower heritability of Whites is attributable to the larger environmental variation in Whites compared with Blacks and not to a smaller amount of genetical variation, since the genetical variation seems to be the same in the two races, given our assumption that A is the same in the two races. We have already discussed the greater E_1 in Whites. However, if E_2 is really larger in Whites than Blacks, then this implies that Whites are subject to a wider range of common environmental effects. Factors producing environmental differences between families may be cultural differences and, therefore, it is possible that there may be a wider cultural range in Whites than Blacks. We can see some evidence for this possibility from the between sub-samples differences in Whites (see Section 1.5.3). Differences in mean between the sub-samples suggests that Whites are not a single population, but that Whites comprise several diverse populations. However, there is no evidence from the means of the sub-samples to suggest that Blacks do not form one homogeneous population. Therefore, there could be greater cultural heterogeneity among Whites than among Blacks explaining the bigger E_2.

However, we must now reconsider the interpretation of these data given that our assumption that A is the same in the two races fails. If A is really larger in Whites than Blacks, we may again account for this by reference to the greater heterogeneity among Whites. Let us assume that Whites comprise many different sub-populations, but that Blacks form one homogeneous population. This is not unreasonable since Whites in America originate from many different countries whose populations are genetically diverse. Mating will, therefore, occur more frequently among members of the same sub-population and this would lead to an increased level of assortative mating in Whites.

5 SUMMARY

We have shown that there are mean differences in "general" ability between Whites and Blacks and that Whites are more variable overall than Blacks. The greater variance in Whites may be partly ascribed to the greater E_1 in Whites. We also have evidence either that E_2 is greater in Whites or that assortative mating is greater in Whites, or else, both E_2 and assortative mating are greater. We have postulated that Whites are genetically and culturally heterogeneous and may be divided into sub-populations on the basis of origin. We have discussed how this diversity among Whites could account for the greater E_2 and/or degree of assortative mating in Whites.

SECTION 8: SUMMARY

Many of the results of this study cannot be generalised to all fourteen measures of ability. Therefore the main analyses and findings will be enumerated, and it is hoped that this summary will enable the reader to determine quickly where to find detailed discussion of points of particular interest.

The work may be summarised:

1. The origination of the Georgia Twin Project by Osborne and Vandenberg in the early sixties is discussed. The location of the subjects through schools and the sample structure are described.
2. Zygosity was determined by a variety of procedures in different parts of the data. These included blood typing, discriminant function analysis of physical measurements and rating by examiners. Errors of diagnosis which might have occurred would lead to predictable biases and the proportion of genetical variation due to genetical causes would be underestimated.
3. The measures of ability were chosen to represent four of Thurstone's (1938) primary mental abilities, with particular weight being given to spatial tests. A factor analysis identified only two factors in these data - one Spatial, one Verbal-Arithmetic. This suggests either that the four factor model of ability is inappropriate and some other model, such as Jensen's two factor model involving Level I and Level II abilities, is a more adequate representation of the true situation, or that the battery of tests chosen does not sample all types of ability equally, and certain types of test are under-represented.

4. Various sources of bias in the data are discussed. These should have only a minor effect on our analyses and should lead to failure of simple models if they are serious.
5. Test score distributions are approximately normal. Mazes errors provides one exception. Analysis is, therefore, performed on the log Mazes errors scores, which are normally distributed.
6. Test score is shown to depend on age. Means and covariances were adjusted to remove the age effect.
7. The usual pattern of sex differences and race differences in mean performance was found. The mean scores of Blacks were lower than those of Whites. Males performed better than females on all Spatial tests. The effects of race, sex and zygosity and their interactions on the means were estimated, by a linear model fitting procedure.
8. Analysis of intraclass correlation coefficients was carried out. This confirmed the inadequa ies of the correlation approach as demonstrated by Eaves and Jinks (1972). A heritable component of variation was demonstrated in most cases, but the power of the test was low and no heritable component could be found for the more unreliable tests. There was no evidence for sex linkage for the Spatial tests. Some evidence for larger correlations in Whites was found for one or two of the Spatial tests. However, a race difference in heritability was not found.
9. The White twins were generally more variable than the Blacks, and empirical and theoretical genotype-environmental models were fitted to the data, in order to compare the components of variation in the two races.

10. A genetical component of variation was essential to explain variation in Verbal-Arithmetic tests, and most Spatial tests, but the power of the tests was not great enough to discriminate between simple genetical and simple environmental models in a few cases. In general, E_1, G_1 and B were all larger in Whites than Blacks. Estimates of the biased broad heritability are slightly larger in Whites than Blacks, but little importance can be attached to this finding because the difference is not significant.

11. Limits on the size of assortative mating and common environmental parameters were set. Only tentative conclusions could be drawn from these because there was much inter-test variation, and race differences were not significant. However, generally either E_2 or assortative mating was larger in Blacks than Whites for Verbal-Arithmetic tests and the reverse was true for Spatial tests.

12. A detailed investigation of the problems of genotype-environment interaction and scale was carried out. A psychometric model of test behaviour was proposed and tested. This model was inadequate in most cases and confirmed the heterogeneous nature of the tests. It was shown that many of the apparent interactions between environmental and genetical differences can be attributed simply to the construction of the tests.

13. A "general factor", extracted by principal component analysis was analysed. The factor scores were normally distributed and there were no problems of scale. Whites were again more variable than Blacks. Environmental models could not explain the observed pattern of variation, and assuming the between families component, B, to be produced entirely by assortative mating, as much as 90% of the

variation could be attributed to genetical differences. Specific environmental variation was larger in Whites than Blacks which could partly explain the greater variability in Whites.

14. Analysis of the between families component, B, showed that either assortative mating or common environmental effects are bigger in Whites than Blacks. There was little evidence for any E_2 in Blacks.

15. In the analysis of individual tests, different contributions of assortative mating or E_2 were found for Verbal-Arithmetic and Spatial tests. Most of the tests are spatial in nature and spatial tests are thought to have the highest loading on 'g'. Therefore, it is not surprising that the results for the general factor are similar to those for the Spatial tests.

16. The greater differences between families in Whites could reflect cultural differences within populations. However, it is suggested that it could be simply due to the greater heterogeneity of the White population in the United States. Analysis of three sub-samples of the data showed that some of the heterogeneity could be attributed to differences in the mean performance of Whites from different regions. We postulated that the division of Whites into partly inbreeding sub-populations on the basis of origin might explain the observed genetical or cultural heterogeneity.

GENETICAL ASPECTS OF HUMAN BEHAVIOUR

A thesis presented to the

Faculty of Science and Engineering

of the

University of Birmingham

by

Krystyna Last

for the degree of

Doctor of Philosophy

Department of Genetics

University of Birmingham

January, 1978

The following has been excluded at the request of the university

Volume 2

Pages 615-635

PART B

A STUDY OF SOCIAL ATTITUDES

INTRODUCTION

The study of the inheritance of intelligence has a long if somewhat controversial history, as we have seen in Part A. In Part B, we turn to two traits which have interested psychologists and sociologists for some time, but have rarely been subjected to the detailed analysis necessary to determine the causes of variation. In the past, most theorists accepted the untested assumption that all attitudes are developed through experience (Chein, 1948; Doob, 1947), more through conviction than overwhelming empirical evidence. However, the study of evolutionary genetics suggests a biological basis for the determination of attitudes and recognition of this may serve to illuminate certain sociological or psychological perspectives in a more realistic light, by suggesting possible selective advantages for the two traits, how they may have evolved within a social structure and possible changes that may occur in the future.

Indirect evidence for a heritable component of variation in attitudes comes from studies of behavioural traits which may lead an individual to prefer one attitude over another. For example, racial or other prejudices often have an element of active hostility and there is abundant evidence for a heredity component of aggression (e.g. Carthy and Ebling, 1964; Scott and Fuller, 1965; Lorenz, 1966). In a recent book, Dawkins (1976) makes the link between behaviour and evolution which is central to the biological understanding of behaviour, by expressing the distinction between motives and consequences in these terms:

"It is important to remember that the above definitions of altruism and selfishness are behavioural, not subjective. I am not going to argue about whether people who behave altruistically are 'really' doing it for secret or subconscious selfish motives. Maybe they are and maybe they aren't, and maybe we can never know..... My definition is concerned only with whether the effect of an act is to lower or raise the survival prospects of the presumed altruist and the survival prospects of the presumed beneficiary".

It is not how we feel or think we are behaving that are of ultimate importance, but the consequences of the actions (i.e. behaviour), prompted by our emotions or attitudes, which are important for our own survival and, through our offspring, of similar feelings and attitudes in the future.

Attitudes are only of evolutionary consequence in so far as they influence survival. Attitudes may be shaped by underlying personality traits, such as level of aggressiveness, which are known to have a genetic basis. This suggests that there may be a genetic predisposition towards certain attitudes. Conversely, particular forms of behaviour are knwon to be under genetical control, and insofar as attitudes are shaping this behaviour, it seems likely that the attitudes themselves have some genetic mechanism. All this suggests that attitudes and personality may have a common biological basis, which is related to survival.

The involvement of a genetic mechanism in determining attitudes does not imply that environmental and cultural factors are unimportant. Behavioural differences may have genetical or social consequences either of which may modify the range of behavioural differences in future generations. A change in social system originated by the behaviour of individuals conditioned by that social system may have consequences for the

genetic system. Conversely, selective pressures changing the gene frequencies may alter the social structure. Thus, any change in a population is the result of both genetical and environmental changes upon the behaviour of its members, and to suggest a genetic basis for the attitudes determining behaviour does not preclude the importance of cultural change. Indeed, a new and important area of study is the cultural transmission of social attitudes through genetical differences, acting environmentally, producing cultural change.

Previous work on the genetics of social attitudes comprises two twin studies, which give a consistent picture for the components of variation which can be separated using twins reared together. The work outlined in this thesis contains an analysis of twin data and confirms the results of earlier studies, but also provides a unique opportunity for the resolution of the cultural components of variation and cultural transmission against a background of simple additive genetical differences and the effects of the mating system.

Eysenck became interested in social attitudes and re-examined three early studies (Thurstone, 1934; Carlson, 1934; Ferguson, 1939). In a series of investigations (Eysenck, 1944, 1947) he showed that social and political actions of all kinds are mediated through attitudes and opinions, which are organised into two principal orthogonal factors. These he labelled Radicalism and Toughmindedness. The development and validation of these two dimensions of attitudes was carried out by Coulter (1953), George (1954) and Melvin (1955) and is summarised by Eysenck (1954), together with his own work. Subsequently, the validity of this theory was demonstrated, since the two dimensions showed a relationship with social class and were sufficient to account for the relationship between the different political parties in this country

(Eysenck, 1951, 1954, 1956). He also showed that the theory could be generalised to populations from the United States, Sweden, Germany and other western countries (Eysenck, 1953; Dator, 1969; Bruni and Eysenck, 1976).

Later, Eysenck showed that the Radicalism-Conservatism dimension could be usefully sub-divided into two components, social conservatism and economic conservatism (the latter being a capitalism-socialism dimension) (Eysenck, 1975, 1976a, b).

A lot of evidence has accumulated in support of the general theory of social attitudes using a variety of questionnaires and types of question (e.g. Rokeach, 1973; Wilson, 1973; Schubert, 1975). Indeed, the data analysed here also confirm the relevance of Radicalism and Toughmindedness and were obtained as responses to one of the most recent types of questionnaire. The most recent study (Hewitt et al, 1977) replicates the findings of the earliest studies after 25 years with remarkable consistency.

Study of the inter-relationship of social attitudes and the whole personality showed that Toughmindedness is related to extraversion and Psychoticism (e.g. Eaves and Eysenck, 1974). In particular there is a distinct tendency for Toughmindedness to be associated with aggressiveness and dominance. Wilson (1975) has suggested personality correlates of Radicalism.

Eaves and Eysenck (1974) also studied the causes of trait variation and covariation in three personality variables and the two dimensions of social attitudes, in order to determine the extent to which variation in personality and opinion share a common genetic basis. They were able to discriminate between simple genetical and simple environmental models and show a heritable component of the observed variation and covariation.

The possibility of common environmental influences was suggested. Genetical differences might account for as much as 65 per cent of variation in Radicalism scores and 54 per cent in Toughmindedness scores. Despite this, parental attitudes will be relatively poor predictors of offspring attitudes, since on the basis of the simple genetical model, given random mating, the parent offspring correlations are expected to be roughly between 0.2 and 0.3. This is substantially lower than correlations reported earlier (e.g. Fuller and Thomson, 1960). Eaves and Eysenck (1974) suggest that if the change is real, it might reflect cultural change over the last 30-40 years and could result from increased social mobility minimising the importance of the family environment, and increasing the importance of genetical differences and environmental experiences unique to the individual. This suggests that attitudes are sensitive to cultural change and may be subject to cultural transmission which is the particular interest of the present work.

Hewitt (1974) analysed the individual dimensions of social attitudes and showed that specific environmental influences, the family environment or the effects of the mating system and additive genetical differences may each account for approximately one third of the total variation in Radicalism. Almost identical results were obtained for the relative contributions of the different sources of variation in a later twin study (Martin, 1976; Martin and Eysenck, 1976). Similarly, both twin studies were consistent with a model for Toughmindedness in which the expression of genes and the effects of the environment are dependent on sex, suggesting that underlying variation in toughmindedness, there is some mechanism of sex limitation mediated genetically or culturally. The

possibility of a cultural component for both traits (demonstrated by Eaves (1977) for Toughmindedness in the earlier twin study), means that some variation in each trait could be produced by the cultural transmission of phenotypic differences. Twin data are unsuitable for the study of cultural transmission since cultural differences and the effects of the mating system are confounded. Thus the earlier studies could only suggest a tempting line of future research. Adoption data is probably the most relevant for the solution of the problem of cultural change and in Part B of this thesis a large body of twin and adoption data are analysed. We hope that this will enable some advance to be made in understanding the inter-relationships of the genetical predisposition of individuals, the different types of environmental factors influencing them and the structure of society and cultural change.

SECTION 1: DESCRIPTION OF THE STUDY

1 THE QUESTIONNAIRE

The structure of the questionnaire used in this study differs radically from that of questionnaires previously used in attempts to elucidate the genetical and environmental components of variation in social attitudes. Eysenck (1957) publishes, in full, an example of the earlier type of questionnaire - the 60 item Public Opinion Inventory - together with its scoring key and a discussion. Questionnaires used in the work described in the Introduction differ from this only in detail. Subjects respond to explicit statements regarding specific attitudes. Wilson and Patterson (1968) describe semantic problems associated with the conventional manner of framing the questions. They suggest that presenting single nouns with which subjects can agree or disagree might overcome these problems. This led to the development of the Wilson-Patterson Attitudes Questionnaire, presented in Table B1. Wilson (1973) found that responses to this questionnaire could be summarised adequately by reference to two major factors, identified as "Radicalism" and "Toughmindedness".

Eysenck performed detailed analysis, using responses to a 68 item Wilson-Patterson Attitudes Questionnaire, of a quota sample of 1442 adult subjects whose age and sex distributions closely followed those of the whole population. He identified a number of meaningful primary factors, resembling those found in previous studies, whose inter-correlations gave rise to two major higher order factors. These were closely similar in content to the Radicalism and Toughmindedness

TABLE B 1 : ITEMS OF THE WILSON-PATTERSON ATTITUDES QUESTIONNAIRE

Subjects were asked to circle "Yes","?", or "No", depending on whether they agree with, are uncertain about, or disagree with the following topics.

1. Death Penalty
2. Evolution Theory
3. School Uniforms
4. Striptease Shows
5. Council Housing
6. Sunday observance
7. Hippies
8. Women's Lib
9. Student Protest
10. Police
11. Surtax
12. Wife-swapping
13. Foreign Aid
14. Pop Music
15. Welfare State
16. Tradition
17. Conscription
18. The Pill
19. Patriotism
20. Modern Art
21. United Nations
22. Self Denial
23. Working Mothers
24. Military Drill
25. Co-Education
26. Law Reform
27. Divine Law
28. Socialism
29. White superiority
30. Cousin Marriage
31. Moral Training
32. Suicide
33. Political Systems
34. Legalised Abortion
35. Empire Building
36. Licensing Laws
37. Space Research
38. Strikes
39. Common Market
40. Computer Music
41. Chastity
42. Royalty
43. Women Judges
44. Capitalism
45. Conventional Clothing
46. Teenage Drivers
47. Apartheid
48. Nudist Camps
49. Church Authority
50. Inheritance Tax
51. Astrology
52. Disarmament
53. Censorship
54. Birching
55. Mixed Marriage
56. Strict Rules
57. Arab Politics
58. Pacifism
59. Law and Order
60. Casual Living
61. Divorce
62. Profit sharing
63. Inborn Conscience
64. Coloured Immigration
65. Bible Truth
66. Trade Unions
67. Supersonic Airliners
68. Liberals

factors previously described using conventional types of questionnaire items. Evidence suggests that these factors are partly determined by heredity and are related, genetically and environmentally, to personality variables (Eaves and Eysenck, 1974; Martin, 1976). Despite the novelty of the Questionnaire used, we hope to identify the same factors as previous workers and to compare our genotype-environment analysis of variation in attitudes in society with similar earlier analyses.

2 THE SUBJECTS

The Wilson-Patterson Attitudes Questionnaire was sent, together with questionnaires on neuroticism, impulsiveness and smoking and drinking behaviour, to twins from the Maudsley Twin Register. This register of twins from the London area has been built up over the years from volunteers who responded to advertisements in the press and journals and have agreed to assist the work of the Institute of Psychiatry (University of London) by occasionally filling in postal questionnaires. The responses to the Wilson-Patterson Attitudes Questionnaire of those twins who returned the forms are the basis of this study.

The present study included 587 pairs of twins from the register whose breakdown by zygosity and sex is given in Table B2. Zygosity was determined on the basis of responses to a questionnaire on physical similarity and mistaken identity in childhood, similar to that described by Kasirel and Eaves (1976). They found an accuracy of 96.1% for their questionnaire method compared with blood-typing, using a sample of 178 pairs of twins and fifteen different blood group systems.

The Attitudes Questionnaire was also sent to individuals who were part of an adoption study. The families, containing adopted individuals, volunteered to take part in research, in response to advertisements in newspapers and women's journals. There are data on 445 individuals who returned the questionnaire. This brings the total number of subjects in the study, including the twin and adoption studies, up to 1619 individuals. Table B3 shows the numbers of each type of individual in the adoption study. They are mainly adopted individuals and adopting

TABLE B 2 : BREAKDOWN OF TWIN SAMPLE BY ZYGOSITY AND SEX

	MZ	DZ	Total Number of Pairs
MALES	83	52	135
FEMALES	233	147	380
OPPOSITE-SEX	-	72	72
TOTAL NUMBER OF PAIRS	316	271	587

KEY: MZ - monozygotic twin pair

DZ - dizygotic twin pair

TABLE B3 : INDIVIDUALS IN THE ADOPTION STUDY

TYPE OF INDIVIDUAL	NUMBER
Adopted son	31
Adopted daughter	239
Adopting father	62
Adopting mother	77
Natural son of adopting parents	12
Natural daughter of adopting parents	13
Second adopted son of adopting parents	9
Second adopted daughter of adopting parents	2
Total number of individuals	445

Various Group Totals

Total number of adopted individuals	281
Total number of parents	139
Total number of natural children	25
Total number of males	114
Total number of females	331

parents, but 25 natural children of these parents also took part. Relationships among these 445 individuals are presented in Table B4. The groups shown are not independent, since an individual may be included in more than one group. This precludes any simple analysis of all the summary statistics based on these groups simultaneously. However, the table does provide a useful summary of the sample structure.

TABLE B 4 : RELATIONSHIPSAMONG INDIVIDUALS OF THE ADOPTION STUDY

TYPE OF RELATIONSHIP	NUMBER OF PAIRS
Spouses	61
Adopting father - adopted son	8
Adopting father - adopted daughter	34
Adopting mother - adopted son	21
Adopting mother - adopted daughter	40
Families including one adopted individual and one or both parents	61
Families including one adopted individual and both parents	32
Families including one adopted individual and one parent only	29
Father - son	3
Father - daughter	5
Mother - son	4
Mother - daughter	3
Families including natural child and one or both parents	13
Families including natural child and both parents	9
Families including natural child and one parent only	4
Adopted child - natural child	10
Families with more than two children reared by same parents	15

3 ADEQUACY OF SAMPLING

This is not a random sample, since the subjects were volunteers. Females are clearly over-represented in both parts of the sample (see Tables B2 and B3). It is thought that the lower socio-economic status groups are under-represented. These biases probably reflect the nature of the journals in which advertisements were placed: many were placed in women's magazines.

There is an excess of monozygotic twins; same-sex dizygotic twins are over-represented compared with opposite-sex twins, since we expect equal numbers of the two types of non-identical pair.

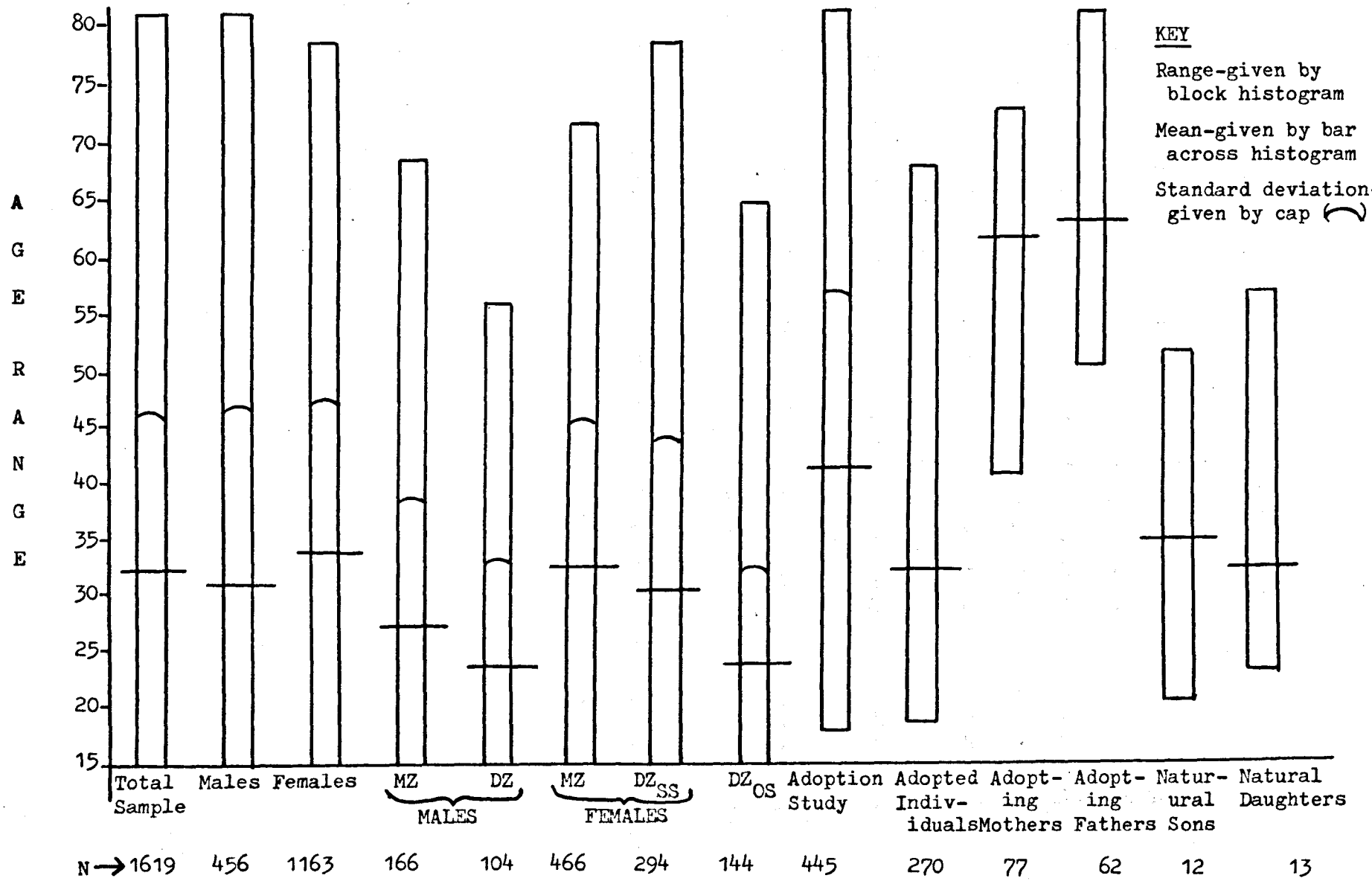
The practical consequences of these biases, for the sampling of Radicalism and Toughmindedness from the population, are not known. However, the sample may not be representative of the population as a whole. There may be a restricted range of genotypes and environments, which could lead to the underestimation of between families effects. This problem will be further discussed when interpreting analyses of the genetical and environmental factors producing variation in attitudes.

4 AGE DISTRIBUTION IN THE SAMPLE

The age range in the sample, from 15 to 82 years, is wide. The mean age is 32.91 $\pm$ 14.88 years. The age distributions in different sub-groups of the data are illustrated in Figure B1. The overall female mean is significantly greater than the male mean ($\chi^2_1 = 6.13$). We notice, also, a tendency for dizygotic twins to be younger than monozygotic twins. The mean ages of the adopted individuals and the twins are similar, but as expected, the adopting parents are much older. Thus, the age structure of the data is such that any relationship between test score and age will make interpretation of the analyses meaningless.

Eysenck (1954) reports evidence that scores on an earlier version of the questionnaire are not age-related. However, we cannot assume that this is true for the present study, since the questionnaire and sample structure are very different. Dependence of attitude score on age will be tested and the appropriate age-corrections made.

FIGURE B1: MEAN AND RANGE OF AGES IN DIFFERENT SUB-GROUPS OF THE DATA



5 ADVANTAGES OF THE DESIGN OF THIS STUDY

The structure of this study has considerable advantages. The large number of twins will enable us to test certain basic models of variation and to compare this study with previous studies. (Eaves and Eysenck, 1974; Martin, 1977). The weakness of the classical twin study of identical and non-identical twins reared together is that it does not allow discrimination between certain hypotheses about the nature of between pairs variation, crucial to our understanding of individual differences in attitudes. The adoption data will enable us to assess the relative importance of cultural effects and assortative mating in producing between families variation and to test the hypothesis of cultural transmission for these traits. Thus, the analysis of individual differences in attitudes will be taken further in this study than has been possible before.

SECTION 2: THE STRUCTURE OF ATTITUDES IN THESE DATA

1 REASONS FOR PERFORMING THE FACTOR ANALYSIS

Previous work with the Wilson-Patterson Attitudes Questionnaire involved large quota samples. Since there may be a restricted range of phenotypes in this study, items may not discriminate adequately between subjects and separation of individuals on the basis of the two major dimensions of Radicalism and Toughmindedness, similar to that achieved in previous studies, may not be possible. Therefore, we will seek evidence that the Radicalism and Toughmindedness scales are appropriate for these data.

2 THE FACTOR ANALYSIS

A general description of the concepts and theory of factor analysis is given in Appendix C. The particular methods used here will be briefly described, but more detail of Principal Factoring can be found in Section 2 of Part A.

Responses of the 1619 individuals to the 68 items of the Wilson-Patterson Attitudes Questionnaire (coded as 0 = Disagree, 1 = Don't know, 2 = Agree) were entered into the analysis. There were no missing observations. All the computations were carried out on the University of Birmingham's ICL 1906A computer using a program from the library of Biomedical Computer Programs (BMD) (Dixon, 1973).

Principal facturing with interaction was used. The advantages and disadvantages of this method are discussed elsewhere. A 68 x 68 matrix of correlations among the items was calculated. Its' leading diagonal elements were replaced by estimates of the communality for each item. A solution for a specific, predetermined number of factors was then sought, and new communalities were calculated from the multiple correlation with the factors. This procedure was iterated until a specified criterion of convergence was achieved. The problem was to decide the appropriate number of factors and the convergence criterion. In this case, convergence was assumed when the maximum change in the communality estimates was less than 0.001, assuming a two factor model. The final communality estimates, which are estimates of the variance of each item shared in common with the other items are given in Table B5 for the two factor model.

TABLE B 5 : FINAL ITEM COMMUNALITIES FOR A TWO FACTOR MODEL

ITEM	COMMUNALITY	ITEM	COMMUNALITIES
1	224	35	099
2	072	36	150
3	210	37	033
4	153	38	287
5	092	39	079
6	202	40	061
7	341	41	212
8	159	42	249
9	345	43	118
10	138	44	074
11	148	45	120
12	213	46	053
13	188	47	114
14	060	48	222
15	111	49	223
16	198	50	114
17	257	51	007
18	076	52	159
19	232	53	148
20	172	54	184
21	098	55	228
22	133	56	232
23	056	57	040
24	204	58	168
25	093	59	128
26	036	60	255
27	256	61	087
28	225	62	107
29	181	63	055
30	069	64	279
31	202	65	257
32	112	66	201
33	051	67	000
34	071	68	101

NOTE: Decimal points are omitted from the communalities.

Two factors were extracted because we wished to test the hypothesis, suggested by previous work, that the two factor model is appropriate for our data . Inspection of the proportion of the total variation accounted for by the first few principal components showed that this was a reasonable procedure. The first principal component accounted for 10.6% of the total variation and the second for 4.4%. The third component accounted for 3.6% of the total variation, and all subsequent components accounted for substantially less than 2%. The dominance of the first component probably reflects the weight given, when constructing the questionnaire, to discriminating between Radicalism and Conservatism.

Factor loadings for the two orthogonal unrotated factors are given in Table B6. Oblique rotation of these factors for simple loadings was performed by attempting to minimise the squared loadings on each factor using an iterative process. This solution of the factor rotation problem has been called direct quartimin. The oblique rotated factor loadings for each item are given in Table B7.

A correlation of 0.199 was found between the two factors. This is fairly high since earlier work suggested that Radicalism and Toughmindedness are orthogonal. However, it is in line with the correlation of 0.18 found by Eysenck in his quota sample. The two factors obtained were easily identified as Radicalism and Toughmindedness from their item content. Thus, despite earlier misgivings about the unrepresentative nature of the sample, these two major dimensions of attitudes, identified by earlier workers, may be appropriate for our data.

TABLE B 6 : UNROTATED ORTHOGONAL FACTOR LOADINGS FOR A TWO FACTOR MODEL

	FACTOR			FACTOR	
ITEM	1	2	ITEM	1	2
1	-388	-271	35	-286	-131
2	256	078	36	-290	256
3	-422	180	37	133	125
4	351	-173	38	529	086
5	194	233	39	018	280
6	-381	238	40	239	064
7	582	051	41	-367	279
8	378	126	42	-445	226
9	565	161	43	053	340
10	-258	268	44	-272	016
11	228	310	45	-318	137
12	352	-299	46	215	084
13	218	375	47	-216	-260
14	244	024	48	463	-085
15	200	267	49	-388	269
16	-329	299	50	294	167
17	-502	-069	51	-058	-057
18	274	039	52	330	225
19	-439	198	53	-329	198
20	396	121	54	-407	-137
21	033	312	55	418	230
22	-194	309	56	-464	129
23	236	028	57	129	153
24	-452	015	58	336	234
25	223	207	59	-260	246
26	053	182	60	483	-148
27	-432	264	61	295	-009
28	440	179	62	162	284
29	-244	-349	63	-161	171
30	262	-020	64	391	355
31	-306	329	65	-437	257
32	319	-101	66	360	268
33	050	220	67	020	-006
34	266	-015	68	064	311

NOTE: Decimal points are omitted from the factor loadings.

TABLE B 7 : OBLIQUE ROTATED FACTOR LOADINGS FOR THE TWO FACTOR MODEL AND COMPARISON WITH LOADINGS OBTAINED BY E SENCK

ITEM		R		T	
		This Study	Eysenck's Study	This Study	Eysenck's Study
1	Death Penalty	-44	-29	-11	-16
2	Evolution Theory	20	42	14	-03
3	School Uniforms	-05	-02	-44	-29
4	Striptease Shows	02	45	39	15
5	Council Housing	30	13	-01	-06
6	Sunday Observance	02	-28	-45	-31
7	Hippies	34	49	41	20
8	Women's Lib	30	35	20	-18
9	Student Protest	43	46	32	08
10	Police	11	03	-38	-30
11	Surtax	39	35	-03	-19
12	Wife-swapping	-09	22	47	40
13	Foreign Aid	44	32	-09	-14
14	Pop Music	14	35	17	15
15	Welfare State	34	18	-03	-12
16	Tradition	10	17	-45	-39
17	Conscription	-32	-18	-34	-25
18	The Pill	17	50	18	-00
19	Patriotism	-05	19	-47	-48
20	Modern Art	31	32	22	07
21	United Nations	29	36	-18	-37
22	Self Denial	18	12	-36	-43
23	Working Mothers	14	21	16	-05
24	Military Drill	-22	-08	-36	-32
25	Co-Education	30	37	03	-13
26	Law Reform	19	33	-08	-14
27	Divine Law	02	-08	-51	-39
28	Socialism	38	21	22	05
29	White superiority	-43	-35	05	13
30	Cousin Marriage	16	24	21	28
31	Moral Training	14	20	-46	-44
32	Suicide	07	26	31	20
33	Political Systems	22	33	-11	-32
34	Legalised Abortion	12	43	21	-04

contd...

TABLE B 7 continued.

ITEM		R		T	
		This Study	Eysenck's Study	This Study	Eysenck's Study
35	Empire Building	-26	-31	-13	-08
36	Licensing Laws	08	-16	-39	-23
37	Space Research	18	47	02	05
38	Strikes	34	41	35	11
39	Common Market	26	35	-17	-20
40	Computer Music	18	28	14	08
41	Chastity	06	04	-47	-45
42	Royalty	-03	-02	-49	-43
43	Women Judges	33	26	-19	-28
44	Capitalism	-12	12	-22	-25
45	Conventional Clothing	-04	05	-34	-21
46	Teenage Drivers	18	35	11	06
47	Apartheid	-34	-28	01	15
48	Nudist Camps	16	54	41	-09
49	Church Authority	04	-15	-48	-31
50	Inheritance Tax	30	19	11	-05
51	Astrology	-08	10	-01	-11
52	Disarmament	37	18	10	03
53	Censorship	01	-19	-39	-38
54	Birching	-33	-28	-22	-18
55	Mixed Marriage	42	53	17	-00
56	Strict Rules	-12	-21	-44	-24
57	Arab Politics	20	19	-00	-26
58	Pacifism	38	23	10	-06
59	Law and Order	09	04	-36	-30
60	Casual Living	11	17	47	47
61	Divorce	14	39	23	04
62	Profit Sharing	33	21	-07	-11
63	Inborn Conscience	07	30	-24	-38
64	Coloured Immigration	51	34	06	01
65	Bible Truth	01	-29	-51	-38
66	Trade Unions	42	24	10	-05
67	Supersonic Airliners	00	26	02	-00
68	Liberals	31	28	-16	-34

Note: Decimal points are omitted from the factor loadings.

3 COMPARISON OF THIS AND AN EARLIER ANALYSIS

For the purpose of comparison, the rotated factor loadings obtained in this study are followed, in Table B7, by those obtained by Eysenck, using the quota sample described in Section 1. Simple inspection shows a striking correspondence between the two sets of factor loadings. Where differences exist, they reflect the ability of the items to discriminate between individuals in the two studies. However, the loadings are so similar on the majority of items, that the organisation of the attitudes tested by the Wilson-Patterson Attitudes Questionnaire along two dimensions broadly similar to those identified by Eysenck (see Section 1) seems to be appropriate in these data.

4 FACTOR SCORES

The scores of each individual on the Radicalism and Toughmindedness dimensions were estimated and printed by the BMD Library program, used to perform the factor analysis. These "factor scores" are used in all the following analyses of individual differences in social attitudes.

SECTION 3: THE SCALE OF MEASUREMENT

1 CHOICE OF A SCALE

The Radicalism (R) and Toughmindedness (T) scores are a function of the individuals' response to each item and the items' factor loading, summed over items. Thus, the R and T scales are in some sense arbitrary, produced as an artefact of the statistical complexities of factor analysis. Psychological or genetical criteria might lead us to transform the scores if this would facilitate analysis and interpretation of the data, without introducing undesirable statistical properties.

The problem of choosing the most appropriate scale of measurement is discussed in general by Mather and Jinks (1971) and in the context of human behaviour by Eaves et al (1977). Several criteria may be employed in seeking the best scale, which simplifies the analysis and interpretation of data and maximises the predictive value of the results.

1.1 Statistical Criteria

Test scores for continuously varying traits, such as Radicalism and Toughmindedness, are expected to follow the normal distribution, approximately at least. Statistical techniques employed in analysing continuous data are based on the normal distribution and may be inappropriate when significant deviations from normality occur. Biases in estimates of the components of variation may be produced and

the power of significance tests is usually reduced. Therefore, on statistical grounds, we should seek scales which are normally distributed. However, normality is of secondary importance, compared with psychological and genetical properties of the scale.

1.2 Psychological and Genetical Criteria

Psychological and genetical criteria for choosing a scale of measurement overlap to a large extent. A useful descriptive model for variation in the trait, which allows prediction about behaviour in other situations, is needed on psychological grounds. We are also looking for a scale which facilitates genetical analysis and interpretation of the data and allows predictions to be made about the degree of similarity of other relatives. In both cases, we need a measure where non-additivity is minimised. Psychologically this means that we are looking for a scale with equal intervals so that the discriminating power of the test is equal in all parts of the scale. Reliability should not vary in different parts of the scale. Genetically, we require that the sensitivity of the organism to its environment should be independent of its genotype (i.e. there should be no genotypes x environment interaction.) Finding a scale which satisfies these criteria and minimises non-additivity will increase the predictive value of our results within an economical model.

2 DISTRIBUTION OF RADICALISM AND TOUGHMINDEDNESS

Two tests for deviations from normality, using third and fourth moments about the mean, were described in detail in Section 2 of Part A. Coefficients of skewness and kurtosis were calculated for the total sample and for various sub-groups of the data, using the Statistical Package for the Social Sciences (SPSS) (Nie et al, 1970). The probabilities of obtaining these coefficients or larger by chance, based on the null hypothesis of normality of the data, were found from Tables in Snedecor and Cochran (1967). The results of these calculations are summarised in Table B8.

There is evidence of negative skewness for Radicalism in the total sample, but no evidence of kurtosis. The negative skewness indicates an excess of individuals below the mean for Radicalism. Two factors are probably contributing to this excess. Eysenck has shown a correlation between conservatism and social class (see Eysenck, 1957). Thus, skewness could be produced by the under-representation of the lower socio-economic status groups. Age effects are probably also contributing to skewness, since the mean age of this sample is high, and it is known that Conservatism increases with age (Eysenck, 1954).

If we consider the sub-groups, skewness is only highly significant where we have pooled over two or more basic groups, although there is some evidence for skewness, significant at the five per cent level, in MZ males and DZ females. This suggests that the overall skewness may have been largely produced by pooling over heterogeneous groups. Since statistics will be calculated separately for each sub-group, most of

TABLE B 8 : COEFFICIENTS OF SKEWNESS AND KURTOSIS FOR THE RAW FACTOR SCORES

		RADICALISM		TOUGHMINDEDNESS	
		Skewness	Kurtosis	Skewness	Kurtosis
Total Sample	1619	-0.331 **	0.0202 NS	0.355 **	0.605 **
All Females	1162	-0.292 **	0.183 NS	0.750 **	0.837 **
All Males	457	-0.428 **	0.222 NS	0.252 *	-0.097 NS
MZ Males	166	-0.341 *	-0.334 NS	0.416 *	0.097 NS
MZ Females	466	-0.178 NS	0.033 NS	0.740 **	0.627 *
DZ Males	104	-0.050 NS	-0.582 NS	0.109 NS	-0.336 NS
DZ Females	294	-0.258 *	-0.275 NS	0.589 **	0.272 NS
DZ Opposite-sex	144	-0.213 NS	-0.588 *	0.343 *	-0.119 NS
Adopted individuals	270	-0.800 **	1.834 **	0.979 **	1.612 **
Adopting mothers	77	-0.303 NS	0.181 NS	2.248 **	7.616 **
Adopting fathers	62	0.014 NS	-0.429 NS	1.077 **	1.387 *
Natural sons	12	-0.355 NS	-1.171 **	0.542 NS	-1.155 **
Natural daughters	13	0.075 NS	-1.581 **	0.437 NS	-1.178 **

KEY : * Significant at the 5% level
** Significant at the 1% level
NS Not significant

which show no significant skewness, the observed deviation from normality in the total sample may not present problems in the analysis of Radicalism. The observed kurtosis in one or two groups is not reflected by kurtosis in the total sample and should not seriously bias our analyses.

For Toughmindedness, there is highly significant skewness and kurtosis in the total sample and evidence for non-normality in nearly every sub-group. Problems produced by non-normality are likely to be far more serious than for Radicalism.

These deviations from normality could reflect the underlying non-normality of the distributions of genetical or environmental effects. However, variation between "item difficulties" or else sampling bias are more likely explanations.

3 INTERNAL STRUCTURE OF THE SCALES

Jinks and Fulker (1970) propose scaling tests for the detection of systematic genotype-environment interaction, employing third degree statistics. Interactions between the genotype and within families variation can be detected using identical twins reared together. Data on MZ twins reared apart are required to detect similar interactions with the between families environmental component (E_2). These GE_2 interactions, if present, will be confounded with estimates of genetical and environmental components of variation, since they cannot be detected in these data.

Interaction of the between families genetical component (G_2) and E_2 will be confounded with the between families effect. Interaction of within families genetical differences (G_1) and E_2 will be confounded with the G_1 component (Eaves, 1976).

The scaling test of Jinks and Fulker (1970) not only detects systematic G x E, but also demonstrates psychometric inadequacies of the scale, since it uses the regression of pair sums onto absolute pair differences. Differences within pairs of MZ twins are produced by both unreliability of measurement and specific environmental differences (E_1). Therefore, any systematic relationship of pair differences with the pair sums (which provide a measure of the genotype and the common family environment) shows that either the reliability varies in some systematic way in different parts of the scale, or that the magnitude of E_1 depends on the genotype of the individual it is affecting (i.e. there is G x E interaction). Thus, this scaling test provides a test of both genetical and psychological criteria for the adequacy of the scale simultaneously.

Further discussion of the test may be found in Section 6 of Part A, together with details of the methods. Pair sums and absolute paid differences were calculated for all the same-sex twins. Pair sums were regressed onto pair differences using a program from the BMD library. The significance of any linear or quadratic components was tested. Higher order interactions were not calculated since these would be difficult to interpret biologically and psychologically. The analysis was carried out separately for all MZ twins and then separately for males and females. These analyses were repeated for the DZ twins for reasons which will become clear later. The outcome of these analyses is summarised in Table B9.

No significant quadratic regressions were found. A linear component, significant at the 5% level only, was found for Radicalism in the MZ twins as a whole. But strong evidence for a linear component was found in all groups, including the DZ twin groups, for Toughmindedness.

Since there was a relationship between within pairs differences and pair sums in MZ twins, we also expect a regression in DZ twins. The test was made for DZ twins, because a regression in the DZ twins, with no corresponding regression in the MZ twins, would be of great interest. The absence of a regression in the MZ twins would show that the DZ regression was not produced by "G x E" interaction. Therefore, a DZ regression would be produced by the regression of pair sum on genetical differences within pairs. If genetical differences within DZ pairs depend on the scale, then non-additive genetical effects such as dominance and epistasis or unequal gene frequencies must be responsible.

TABLE B 9 : SUMMARY OF SUM-DIFFERENCE REGRESSIONS FOR RAW FACTOR SCORES

	RADICALISM		TOUGHMINDEDNESS	
	Linear	Quadratic	Linear	Quadratic
All monozygotic twins	*	-	***	-
Monozygotic males	-	-	*	-
Monozygotic females	-	-	**	-
All dizygotic twins	-	-	**	-
Dizygotic males	-	-	*	-
Dizygotic females	-	-	*	-

KEY :

* Significant at the 5% level

** Significant at the 1% level

*** Significant at the 0.1% level

\- No significant regression

Since we found no sum-difference regression in DZ twins, there is no evidence for genetical non-additivity in these data. Non-additivity is not precluded, however, since the power of this test is probably low and only a substantial amount of non-additivity could be detected using this test or any other test based on first degree collateral relatives.

In summary, there is little evidence of scalar problems for Radicalism. The slight linear regression in MZ twins taken as a whole might be produced by pooling over sexes. However, absolute differences within pairs increase significantly with Toughmindedness score. Either unreliability increases with Toughmindedness score, or there is genuine G x E interaction, such that the development of Toughmindedness is more susceptible to specific environmental influences. Both alternatives are plausible and we are unable to discriminate between them. Whichever is true, analysis of Toughmindedness on this scale might produce results which are difficult to interpret and require many parameters to procure satisfactory predictions.

These conclusions are similar to those based on the distributions of Radicalism and Toughmindedness scores. Therefore, we shall analyse the raw Radicalism scale. But since the Toughmindedness scale is clearly inadequate, we shall seek a new scale of measurement for this trait, which minimises the problems of the present scale. We shall do this by making a suitable transformation of the raw scores.

4 TRANSFORMATION OF THE DATA

Logarithms and square roots of the factor scores were calculated. These transformations are commonly employed with this type of data (Snedecor and Cochran, 1967). Tests for deviations from normality were performed on the transformed data and the results are summarised in Table B10. Transformation has produced non-normal distributions for Radicalism, suggesting that the original scale is more appropriate. For Toughmindedness, the logarithmic transformation produces no improvement over the original scale. However, on the square root scale deviations from normality have largely disappeared. Skewness for males as a whole may be due to pooling data over heterogeneous groups of males. Generally, however, the distribution of test scores on the square root scale is more satisfactory than on the original scale.

Pair sums were regressed onto absolute pair differences in order to determine whether transformation had changed non-additivity in the scales. The results are briefly summarised in Table B11. Both transformations produced sum-difference regressions for Radicalism. A regression remains for Toughmindedness on the logarithmic scale, but the square root transformation has removed the relationship between pair sums and absolute pair differences.

All further analysis will therefore be based on the untransformed Radicalism scores and square roots of the Toughmindedness scores, since non-normality and non-additivity were minimised on these scales.

TABLE B 10: SUMMARY OF TESTS OF NORMALITY FOR THE TRANSFORMED SCALES

	TOUGHMINDEDNESS				RADICALISM			
	Square root		Logarithm		Square root		Logarithm	
	S	K	S	K	S	K	S	K
Total sample	NS	NS	***	***	***	***	***	***
All males	***	NS	***	***	***	***	***	***
All females	NS	NS	***	NS	***	***	***	***
MZ Males	NS	NS	***	***	***	NS	***	NS
MZ Females	NS	NS	***	NS	***	***	***	***
DZ Males	NS	NS	***	NS	NS	NS	*	NS
DZ Females	NS	NS	***	NS	***	NS	***	***
DZ Opposite-sex	NS	NS	***	*	*	NS	***	NS
Adoption study	NS	NS	***	NS	***	***	***	***

KEY : S - skewness (negative in all cases where significant)
K - kurtosis (positive in all cases where significant)
NS - not significant
* - significant at the 5% level
*** - significant at the 1% level

TABLE B 11 : SUMMARY OF SUM-DIFFERENCE REGRESSIONS FOR TRANSFORMED SCALES

	TOUGHMINDEDNESS				RADICALISM			
	Square Root		Logarithm		Square Root		Logarithm	
	L	Q	L	Q	L	Q	L	Q
All MZ Twins	-	-	***	-	***	-	***	-
MZ Males	-	-	-	-	**	-	***	-
MZ Females	-	-	***	-	***	-	***	-
All DZ Twins	-	-	**	*	*	-	***	-
DZ Males	-	-	-	*	-	-	-	-
DZ Females	-	-	**	-	**	-	***	-

KEY : L - linear component of regression
Q - quadratic component of regression
- no significant regression
* regression significant at the 5% level
** regression significant at the 1% level
*** regression significant at the 0.1% level

SECTION 4: AGE EFFECTS IN THESE DATA

1 INTRODUCTION

Since we have a wide age range in these data and the mean ages of the sub-groups differ, any covariation between attitudes score and age may bias our analyses. A regression of test score on age within groups coupled with a mean difference in age between groups, will produce a mean difference in test score between groups. Significance tests for differences in raw score between groups will, therefore, be meaningless. Thus, we must correct the test scores for any significant relationship with age. This will eliminate one source of bias in our analyses and facilitate interpretation of the data.

2 ANALYSIS OF COVARIANCE

We will consider differences in mean age between groups and their effect on mean differences in Radicalism and Toughmindedness scores, using the analysis of covariance. This statistical technique tests the significance of mean differences in age and test score between groups and also the significance of any regression of test score on age. The means may then be adjusted to remove the age effect and re-tested to see whether any mean differences in test score remain after the age adjustment. The analysis of covariance, thus, combines features of both regression analysis and the analysis of variance. It is discussed by Snedecor and Cochran (1967). Details of the calculations actually performed were described in Section 2 of Part A. We shall confine ourselves to presenting the results of the analysis and discussing their interpretation. The analysis was based on six groups: MZ males, MZ females, DZ males, DZ females, DZ opposite-sex pairs and the individuals of the Adoption Study. The results of this analysis are outlined in the top part of Table B12, (labelled: "1. BASED ON ALL GROUPS"). Since the Adoption Study included both adopted individuals and their adopting parents, the age range was wider than in other more restricted groups and the analysis was repeated using the twin groups only. The results are summarised in the bottom part of Table B12.

Mean differences in age between groups were significant. The Adoption Study included both adopted individuals and their adopting parents. The mean age ($\bar{x}_{adopted} = 41.569$) was much higher than that

TABLE B 12 : SUMMARY OF RESULTS OF AN ANALYSIS OF COVARIANCE

	R	T
1. BASED ON ALL GROUPS		
Significance of differences in mean age	***	***
Significance of differences in mean test score	***	***
Significance of differences in mean test score after age-correction	†	***
2. BASED ON TWIN GROUPS		
Significance of differences in mean age	***	***
Significance of differences in mean test score	*	***
Significance of differences in mean test score after age-correction	†	***

NOTES

R - Radicalism

T - Toughmindedness

*** Significant at the 0.1% level

* Significant at the 5% level

† Significant at the 10% level

ALL GROUPS INCLUDED : MZ males, MZ females, DZ males, DZ females, DZ opposite-sex pairs and the adoption study (i.e. 6 groups)

TWIN GROUPS INCLUDED : MZ males, MZ females, DZ males, DZ females and DZ opposite-sex pairs.

of the twins ($\bar{x}_{twins}$ = 29.625). The significant mean differences in age might be attributable to this group alone. However, significant mean differences in age persisted when the group comprising the Adoption Study was eliminated from the analysis.

In order to determine the source of the observed mean differences in age between the sub-groups of our sample, some simple comparisons were made amongst group means. The female mean was higher than the male mean (t_{1617} = 2.682 $p < 0.01$) and the MZ twin mean was higher than that of the DZ twins (t_{1028} = 2.719, $p < 0.01$).

The female mean ($\bar{x}_{females}$ = 33.538) could be greater than the male mean ($\bar{x}_{males}$ = 31.216) because the mean lifespan of females is greater than that of males. However, samples such as this are not likely to be reflecting population trends, especially since the difference in mean between males and females (2322 years) is rather big. It is more likely that non-random sampling is introducing a bias. Many individuals volunteered to take part in the study through advertisements in newspapers and women's magazines. The mean age of the readership of the type of magazine in which the advertisements appeared is likely to be higher than that of all women between the age of 18 and 82. This would produce a bias towards older women in the sample. In the case of men replying to newspaper advertisements, the age range was not so likely to be restricted.

It is more difficult to see why monozygotic twins were significantly older than dizygotic twins, but it is almost certainly due to biased sampling. We would suggest that this mean difference may be due to the sex difference in mean age. There are a lot more female monozygotic twins than female dizygotic twins in the sample. This would produce a higher

mean age in monozygotic twins as a function of the sex difference.

Significant mean differences in Radicalism and Toughmindedness scores, which might be attributed to the significant mean differences in age between the groups, were found. So the group means were adjusted for the age effect. Highly significant mean differences in Toughmindedness score persisted after age-adjustment. Inspection of the data suggested that these were mainly sex differences, males being more Toughminded than females.

The age-adjustment reduced the significance of mean differences in Radicalism between groups to the ten per cent level. Although these mean differences are not statistically significant, inspection of the means reveals a tendency for females to score less than males on the Radicalism scale.

3 REGRESSION OF ABSOLUTE PAIR DIFFERENCES ON AGE

Any regression of absolute pair differences on age would show a tendency for pair similarity to change with age. For example, in the case of monozygotic twins, a significant, positive linear regression might imply that environmental differences within pairs were making twins less alike as they got older. It could also imply that specific environmental influences were more important in the past producing larger differences between members of the older pairs in the sample. Only a longitudinal study could discriminate between these alternatives. A change in the magnitude of differences within pairs of dizygotic twins with age could be produced by specific environmental effects, but also by genetical differences within pairs. (Eaves and Eysenck, 1976). A regression in DZ twins, unaccompanied by a similar regression in MZ twins, would imply that within families genetical effects differ with age. Genes with different properties might be producing variation at different ages.

In order to see whether there is any evidence for environmental effects altering the degree of similarity between twins, or for different genetical systems operating at different ages, absolute differences between pairs of twins were calculated and regressed onto age using a program from the BMD library of computer programs (Dixon, 1973). The analysis was carried out separately for all monozygotic twins and all dizygotic twins and then repeated for each zygosity-sex group separately.

Only one significant regression was found for Toughmindedness: a quadratic component, significant at the 5% level in MZ males. Specific

environmental influences might make males more dissimilar up to a certain age, after which they become more similar again. It is difficult to visualise a mechanism for this and since the effect was only significant at the 5% level and was not reflected in the DZ males, we shall not consider it further.

Radicalism was more interesting. There was no evidence for a significant regression in the MZ twins, but a quadratic component significant at the 5% level was found in the dizygotic twins taken as a whole and was repeated in both males and females. Although, the regressions were only significant at the 5% level, the consistency over sexes suggests that we may be looking at a real effect.

Since there was no regression in the monozygotic twins, changes in pair similarity with age were probably not produced by specific environmental effects. Therefore, genetical differences within families may be implicated. The pairs are most similar at the age of 15. Differences between members of a pair increase until about 40 years of age, when the twins become more similar again. Two possible mechanisms may be visualised.

i. Developmental rate is the same in both members of the pair i.e. the same operating at a given age. However, they have different alleles of these genes. This produces a difference, at any age, between the individuals. But, at different times in an individual's life, different genes are expressed. If the genes expressed at different times have different non-additive properties (e.g. dominance), then the DZ difference may vary with age as a function of the non-additivity. This seems unlikely for two reasons. Genes controlling developmental rate are

unlikely to be the same in both members of the pair and the non-additive properties of the genes operating would have to change consistently. For example, the dominance of genes being expressed at different times would have to increase up to the age of forty when the maximum DZ difference occurs. After the age of forty, the gradual return to the expression of genes with less non-additivity would have to occur. This situation is illustrated in Figure B2. It implies that the genetic architecture of the trait varies in a consistent manner with age. This is possible but more complex than the second alternative.

ii. This hypothesis suggests that genes producing changes in Radicalism with age are different in the two individuals. Therefore, different genes will be operating in the pair at any time. How this produces the observed relationship between pair difference and age is illustrated in Figure B3.

Both these hypotheses about the origin of the quadratic relationship between absolute pair differences assume that there is a change in Radicalism score with age, and also that this change is probably not linear.

FIGURE B2: CHANGES IN GENETIC ARCHITECTURE WITH AGE

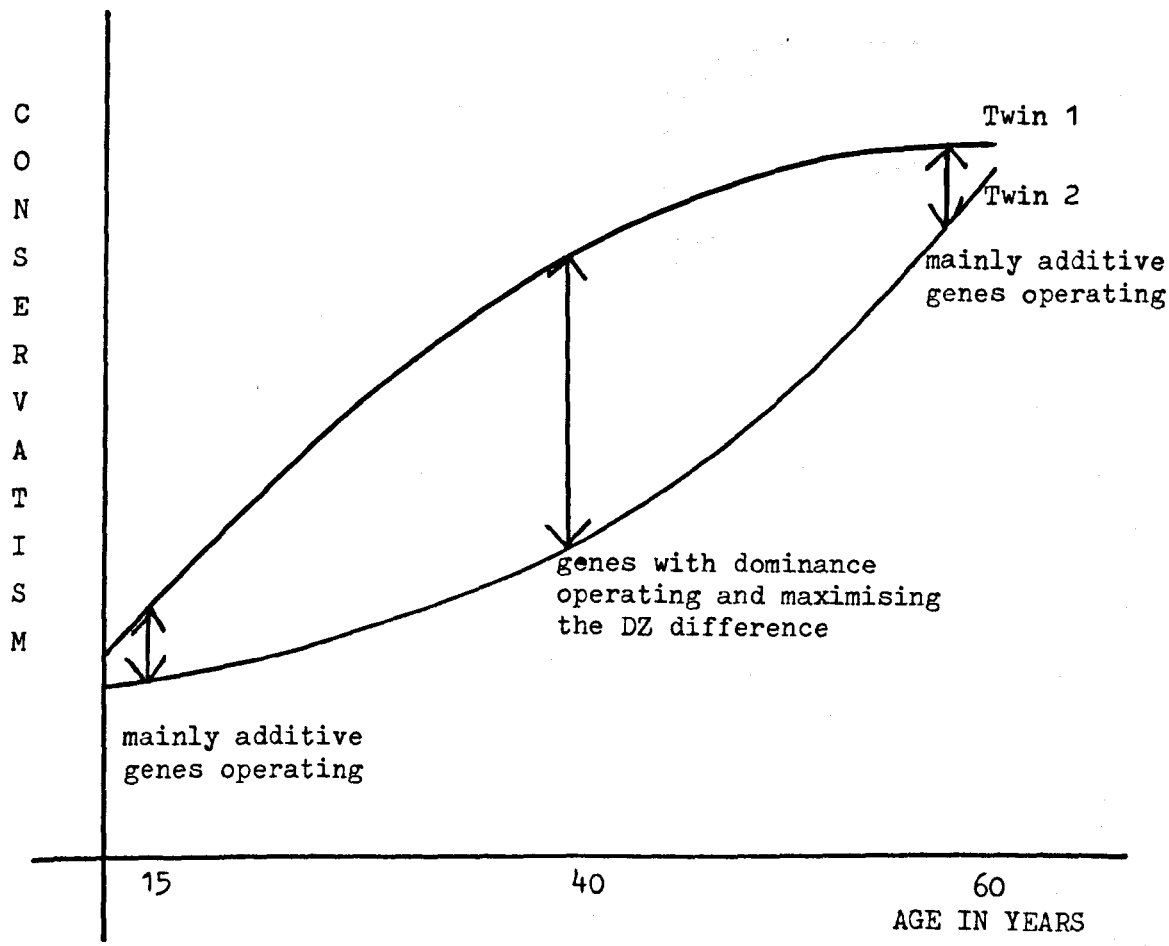
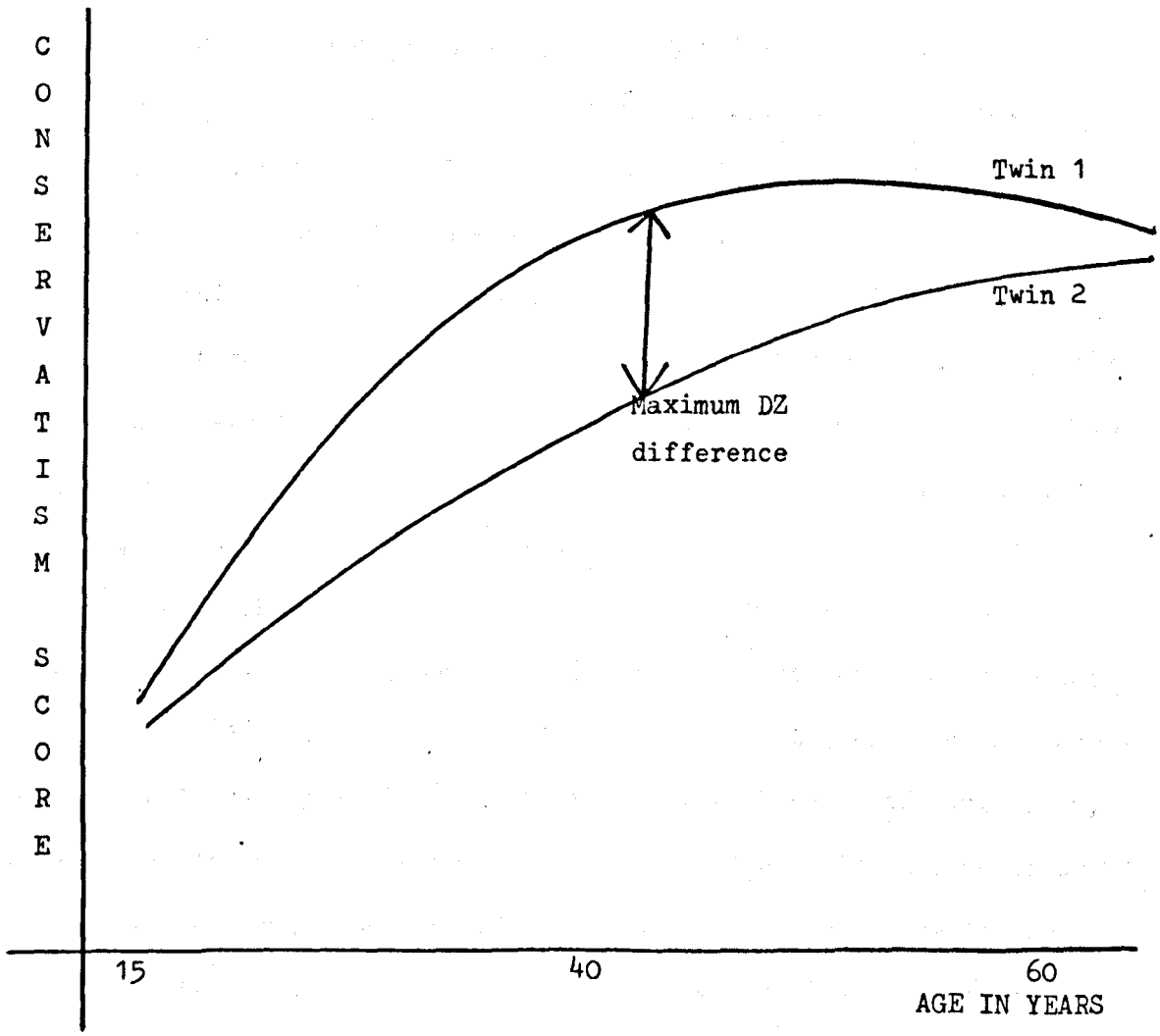


FIGURE B3: CHANGE IN PAIR DIFFERENCE AS A FUNCTION OF DEVELOPMENTAL RATE



NOTE: This graph implies that the increase in Conservatism with age is higher in individuals whose Conservatism score is higher i.e. there is genotypes x ages interaction.

4 REGRESSION OF REAL DIFFERENCES IN OPPOSITE-SEX PAIRS ON AGE

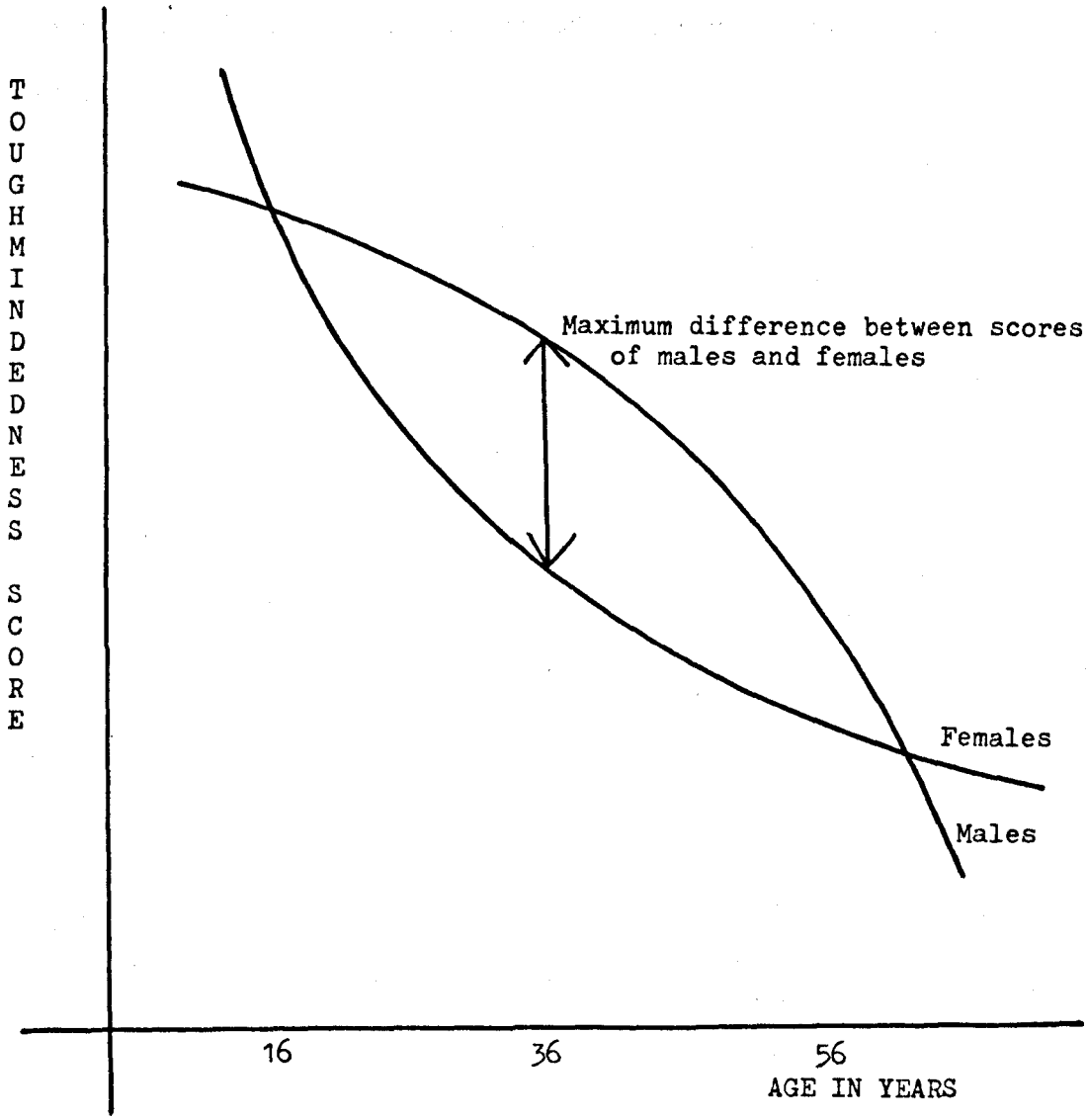
We have already seen that age-adjustment, using the analysis of covariance, does not remove differences in mean Toughmindedness score between sexes. Therefore, the effects of age and sex on test score were examined in more detail, using the opposite-sex pairs. The differences in Radicalism and Toughmindedness scores between males and females were calculated as (Male score - Female score). These differences were then regressed onto pair age.

No significant linear or quadratic component was found for the regression of real differences in Radicalism between members of opposite-sex pairs on age. i.e. the male-female difference does not change with age. This implies that the regression of test score on age will have the same slope in males and females.

Although there was no significant linear component to the regression of differences between members of opposite-sex pairs on age for Toughmindedness, there was a highly significant quadratic component. Males were less toughminded than females below 17 years of age. Then, they gradually become more toughminded compared with females until about 36-40 years of age, when the difference between the sexes begins to decline until, at about 53 years, males are again less Toughminded than females. Thus, males are more Toughminded than females through adulthood, but are less Toughminded in youth and old age.

This regression in opposite-sex pairs leads us to postulate different regressions of Toughmindedness score on age in the two sexes, which are likely to be non-linear (see Figure B4). This is in fact

FIGURE B4: REGRESSION OF TEST SCORE ON AGE IN MALES AND FEMALES, PREDICTED ON THE BASIS OF THE OPPOSITE-SEX PAIRS



what we find in a later section. Again, we have found an interesting relationship between age, test score and sex for Toughmindedness. We shall now attempt to clarify the relationships between these three effects, by performing an analysis of variance of the opposite-sex pairs.

5 ANALYSIS OF VARIANCE OF THE OPPOSITE-SEX PAIRS

The total variation (143 d.f.) in the opposite-sex pairs may be partitioned into the between pairs (71 d.f.) and the within-pairs (72 d.f.) mean squares. The between pairs mean square may be further portioned into the variation between pairs produced by the linear regression of score on age (1 d.f.) and the remaining variation between pairs (70 d.f.) The within pairs mean square was partitioned into variation within pairs produced by the sex difference (1 d.f.), by the interaction between age and sex (1 d.f.) and the interaction between pairs and sex (70 d.f.). This analysis of variance is given in Table B13.

Significant variation between pairs was found for Radicalism and Toughmindedness, but the age effect does not reach significance in the opposite-sex pairs (although later we fill find a significant regression of test score on age in the data taken as a whole).

Mean differences between the sexes were removed by age-adjustment in the analysis of covariance for Radicalism, but not Toughmindedness. But in this analysis of variance the sex effect is not significant for either trait, although we might expect it to be significant for Toughmindedness. The sexxage interaction mean square which is obtained as the linear regression of real pair differences on age also fails to reach significance. We know, however, from the previous section that there is a significant quadratic component to the regression of real pair differences on age, implying a complex interaction of age and sex for Toughmindedness and suggesting that we may have to perform age-correction of the factor scores separately by sex for this trait.

TABLE B 13 : ANALYSIS OF VARIANCE OF THE OPPOSITE-SEX PAIRS

		RADICALISM			TOUGHMINDEDNESS		
ITEM	DF	SUM OF SQUARES	MEAN SQUARE	F	SUM OF SQUARES	MEAN SQUARE	F
BETWEEN PAIRS	71	172.25	2.43	2.48**	15.91	0.22	2.44**
Age	1	4.71	4.71	1.97NS	0.54	0.54	2.45NS
"Pairs" (error)	70	167.55	2.39		15.37	0.22	
WITHIN PAIRS	72	70.84	0.98		6.49	0.09	
Sex Difference	1	0.99	0.99	1.02NS	0.03	0.03	0.33NS
Sex x Age	1	1.17	1.17	1.21NS	0.00	0.00	0.01NS
pairs x Sex	70	67.70	0.97		6.43	0.09	
TOTAL	143						

KEY : DF degrees of freedom
NS not significant at the 5% level
* significant at the 5% level
** significant at the 1% level

6 REGRESSION OF TEST SCORE ON AGE

6.1 Regression of Radicalism and Toughmindedness Scores on Age

Radicalism and Toughmindedness scores were regressed onto age separately for the six sub-groups of the data using a program from the BMD library of computer programs and taking out the linear and quadratic components. The results are summarised in Table B14. The major trend in both Radicalism and Toughmindedness is a negative linear component to the regression. This means that older individuals are less Radical and less Toughminded in these data. Therefore, either conservatism and tendermindedness increase with age or else radicalism and toughmindedness have become more dominant in society over the last 60-70 years. Discrimination between these alternative hypotheses about the origin of the regression of test scores on age is not possible without a longitudinal study.

In the two smallest groups (DZ males and DZ opposite-sex pairs), the regression of Radicalism score does not differ significantly from zero. However, we have evidence to suggest that there is a strong linear regression of Radicalism score on age in the data as a whole. A negative quadratic component was significant at the 5% level in two groups (MZ males and DZ females). Visual inspection of the six regression curves suggested that the rate of decline in Radicalism was somewhat faster after the age of 40. For the purposes of age correction we shall ignore this small quadratic component of the regression.

TABLE B 14 : REGRESSION OF TEST SCORES ON AGE

		RADICALISM		TOUGHMINDEDNESS	
		REGRESSION COEFFICIENT OF LINEAR REGRESSION	SIGNIFICANCE LEVEL OF QUADRATIC COMPONENT	REGRESSION COEFFICIENT OF LINEAR REGRESSION	SIGNIFICANCE LEVEL OF QUADRATIC COMPONENT
MZ MALES	166	-0.0139 *	*	-0.0038 *	NS
MZ FEMALES	466	-0.0105 ***	NS	-0.0089 ***	NS
DZ MALES	104	0.0025 NS	NS	-0.0074 *	NS
DZ FEMALES	294	-0.0153 ***	*	-0.0077 ***	NS
DZ OPPOSITE-SEX	144	0.0145 †	NS	-0.0046 †	***
ADOPTION DATA	445	-0.0113 ***	NS	-0.0098 ***	NS

KEY :

† significant at the 10% level

* significant at the 5% level

** significant at the 1% level

*** significant at the 0.1% level

The regression of Toughmindedness on age was significant in all groups, apart from the opposite-sex pairs, where the probability was still less than 10%. However, in the opposite-sex pairs there was a highly significant quadratic component. Visual inspection of the six curves showed that the regressions were different in the two sexes as predicted in Section 4.4. There was a tendency towards a negative quadratic component in males and a positive quadratic component in females, neither of which was significant. Together, these produce the highly significant quadratic component observed in the opposite-sex pairs. We expected to find results like this when regressing Toughmindedness score on age from the regression of differences between members of opposite-sex pairs on age.

6.2 Homogeneity of the Regression Lines

The results of the regression analysis lead us to suspect that the regressions found in the six sub-groups of the data may not be homogeneous. Assuming linearity, the homogeneity of the residual variances, the slopes and the elevations of the six regression lines was tested, as described in Snedecor and Cochran (1967). The residual variances and slopes were homogeneous for both Radicalism and Toughmindedness. However, differences between the elevations were significant for Toughmindedness ($p < 1\%$) and approached significance for Radicalism ($5\% < p < 10\%$). This means that the regression of scores on age are similar in all groups, but that the means differ. Significant mean differences in Toughmindedness score may be largely attributed to a sex difference. Similarly males tend to be more

Radical than females although mean differences do not reach significance. Because of these mean differences we will correct the data for the regression of test score on age separately by sex, for both traits.

7 AGE-CORRECTION OF THE FACTOR SCORES

In order to perform the age-correction, we must obtain regression coefficients for all males and for all females and find the deviation scores (i.e. the difference between the observed score and the score predicted from the regression of test score on age). These scores are the age-corrected factor scores, with males and females varying around their own mean.

The calculations performed are summarised in Table B15, in which the regression analyses and the regression equations obtained from them are given. Predicted scores for each individual were found using these equations, and hence the deviation scores were obtained. These scores are used in all further analyses.

TABLE B15 : REGRESSION OF RADICALISM AND TOUGHMINDEDNESS SCORES ON AGE FOR MALES AND FEMALES

RADICALISM

	MALES	Y = 0.302 - 0.008X				FEMALES	Y = 0.444 - 0.013 X			
ITEM	DF	SUM OF SQUARES	MEAN SQUARE	F	P	DF	SUM OF SQUARES	MEAN SQUARE	F	P
Regression	1	8.2569	8.2569	8.98	0.003	1	42.7832	42.7823	56.59	0.000
Residual	451	414.8016	0.9197			1160	877.0118	0.7560		

TOUGHMINDEDNESS

	MALES	Y = 1.865 - 0.009X				FEMALES	Y = 1.722 - 0.008X			
ITEM	DF	SUM OF SQUARES	MEAN SQUARE	F	P	DF	SUM OF SQUARES	MEAN SQUARE	F	P
	1	9.2078	9.2078	117.39	0.000	1	16.8979	16.8979	223.24	0.000
	451	35.3738	0.0784			1160	87.8064	0.0757		

SECTION 5: PRELIMINARY ANALYSIS

1 THE MEANS

Means of the age-corrected Radicalism and Toughmindedness scores were calculated for the total sample and for various sub-groups of the data and are presented in Tables B16 and B17. Factor scores were calculated around a mean of zero and subsequently sexes were age-corrected separately around their own mean. Thus all the means are close to zero. However, the overall means of males and females are not exactly zero because they were based on all 1619 individuals, whereas the age-correction was performed on only the 1615 individuals for whom the ages were known.

A simple linear model, specifying the effects of zygosity, sex and their interaction (see Table B18) was fitted to the means by weighted least squares, using the reciprocals of the variances of the means as weights. The procedure is described in Section 3 and Appendix E. The estimates of the effects of zygosity, sex and their interaction on the means were not significant for either Radicalism or Toughmindedness. Since these scores for each sex were adjusted around a mean of zero, lack of significant sex differences between means is no surprise and is not very informative. The absence of zygosity difference suggests no differential sampling bias of the twins. Therefore, we went back to the unagecorrected scores and fitted the linear model to the raw Radicalism factor scores and the square roots of the factor scores for Toughmindedness. The results are presented in Table B18.

TABLE B 16 : MEANS OF AGE-CORRECTED RADICALISM SCORES

	MALES (M)	FEMALES (F)	M + F
M^Z TWINS	0.119	-0.013	0.021
SAME-SEX D^Z TWINS	-0.121	0.016	-0.019
OPPOSITE-SEX D^Z TWINS	0.095	0.137	0.116
ALL TWINS	0.027	-0.002	0.019
All individuals except adopted individuals	-0.003	-0.004	0.012
ADOPTED INDIVIDUALS	-0.292	-0.027	-0.057
ADOPTING PARENTS	-0.115	-0.060	-0.084
NATURAL CHILDREN OF ADOPTING PARENTS	0.497	0.168	0.326
ALL ADOPTION STUDY	-0.104	-0.031	-0.050
ALL DATA	0.005	-0.002	0.000

TABLE B 17 : MEANS OF AGE-CORRECTED, TRANSFORMED TOUGHMINDEDNESS SCORES

	MALES (M)	FEMALES (F)	M + F
MZ TWINS	0.015	-0.026	-0.015
SAME-SEX DZ TWINS	-0.027	0.016	-0.017
OPPOSITE-SEX DZ TWINS	0.013	0.044	0.028
ALL TWINS	-0.001	-0.021	-0.011
ALL INDIVIDUALS EXCEPT ADOPTED INDIVIDUALS	0.022	0.002	-0.008
ADOPTED INDIVIDUALS	0.072	0.059	0.061
ADOPTING PARENTS	-0.048	0.004	-0.019
NATURAL CHILDREN OF ADOPTING PARENTS	0.107	0.019	0.061
ALL ADOPTION STUDY	0.036	0.043	0.041
ALL DATA	0.010	0.001	0.004

TABLE B 18 : SOURCES OF DIFFERENCES IN THE MEANS OF THE TWINS

		MEAN	ZYGOSITY (Z)	SEX (S)	Z X S
Monozygotic Pairs	Males	1	2	1	2
	Females	1	2	-1	-2
Dizygotic same-sex pairs	Males	1	-1	1	-1
	Females	1	-1	-1	1
Dizygotic opposite-sex pairs	Males	1	-1	1	-1
	Females	1	-1	-1	1
RADICALISM					
	Estimate	0.0578	0.0139	0.0079	0.0478
	Significance level	*	NS	NS	**
TOUGHMINDEDNESS					
	Estimate	1.5421	-0.0071	0.0895	0.0087
	Significance level	***	NS	***	NS

There was no significant effect of zygosity on the means for either trait. This shows that being an identical twin rather than a non-identical twin has no effect on attitudes. Although in the total sample males were more Radical than females, the effect of sex on the twin means was not significant. However, there was a sex x zygosities interaction significant at the 5% level. Monozygotic males and dizygotic females were more Radical. There is no obvious explanation for this.

There was a highly significant mean difference between the sexes for Toughmindedness, females being more tenderminded. This confirms the importance of the role of sex for Toughmindedness, which we have already seen during the age-correction, and suggests that this study may support the results of previous workers (Eaves, 1977; Martin, 1977), who found different mechanisms of variation in the two sexes.

2 THE VARIANCES

2.1 Homogeneity of the Total Variances

Total variances were calculated for age-corrected Radicalism and Toughmindedness scores in the total sample and in various sub-groups of the data. These variances are given in Tables B19 and B20.

Bartlett's test was used to test the homogeneity of the total variances of the main sub-groups i.e. 5 twin groups (MZ males and females, DZ males and females and DZ opposite-sex pairs) and 6 groups from the Adoption Study (adopted males and females, adopting mothers and fathers and natural sons and daughters of the adopting parents) Since these variances were homogeneous for both Radicalism and Toughmindedness, we may regard all types of individual as belonging to one population and attempt to explain variation in all groups simultaneously.

However, more powerful tests may be made by comparison of groups of direct interest. For example, comparison of the total variance of all females with that of all males is significant at the 5% level for Radicalism ($\chi^2_1 = 6.24$) and at the 1% level for Toughmindedness ($\chi^2_1 = 6.70$). This indicates some heterogeneity between the sexes, which could lead to problems in attempting to explain variation in males and females simultaneously.

Eaves (1976a) has shown how competition or co-operation between sibs can lead to a distinct pattern of inequality in the total variances of MZ twins, DZ twins and singletons. For competition we expect:

TABLE B 19 : TOTAL VARIANCES OF AGE-CORRECTED RADICALISM SCORES

	MALES (M)	FEMALES (F)	M + F
MZ TWINS	0.716	0.688	0.698
SAME-SEX DZ TWINS	0.845	0.804	0.816
OPPOSITE-SEX DZ TWINS	1.023	0.755	0.884
ALL TWINS	0.777	0.732	0.761
ALL INDIVIDUALS EXCEPT ADOPTED INDIVIDUALS	1.008	1.038	0.794
ADOPTED INDIVIDUALS	1.672	0.736	0.845
ADOPTING PARENTS	0.971	1.055	1.011
NATURAL CHILDREN OF ADOPTING PARENTS	0.599	1.043	0.824
ALL ADOPTION STUDY	1.189	0.817	0.911
ALL DATA	0.919	0.758	0.803

TABLE B 20 : TOTAL VARIANCES OF TRANSFORMED, AGE-CORRECTED TOUGHMINDEDNESS SCORES

	MALES (M)	FEMALES (F)	M + F
MZ TWINS	0.073	0.078	0.077
SAME-SEX DZ TWINS	0.080	0.067	0.070
OPPOSITE-SEX DZ TWINS	0.081	0.069	0.075
ALL TWINS	0.075	0.074	0.075
ALL INDIVIDUALS EXCEPT ADOPTED INDIVIDUALS	0.173	0.087	0.081
ADOPTED INDIVIDUALS	0.075	0.082	0.081
ADOPTING PARENTS	0.125	0.088	0.104
NATURAL CHILDREN OF ADOPTING PARENTS	0.041	0.098	0.069
ALL ADOPTION STUDY	0.146	0.084	0.100
ALL DATA	0.094	0.077	0.082

MZ < DZ < Singletons and for co-operation: MZ > DZ > Singletons. However, we do not observe either pattern, and the Bartlett test gives no reason to suppose that the total variances are heterogeneous. Thus, there is no evidence that the phenotype of one sib acts environmentally on that of the other, making it either more or less similar to itself.

A priori, it seems likely that cultural effects may be important in determining attitudes and cultural transmission could play a role in producing variation. Eaves (1976b) has shown that when a trait is subject to cultural transmission, the variance of adopted individuals is expected to be greater than that of individuals reared by their natural parents. In these data, the total variance of adopted individuals equals that of individuals reared by their natural parents for Toughmindedness, providing no suggestion of cultural transmission. Variation in Radicalism of the adopted individuals is non-significantly greater than that of individuals reared by their natural parents ($\chi^2_1 = 0.44$, $0.70 > p > 0.50$). Thus, although cultural effects could be important in determining individual differences in attitudes, the total variances provide no evidence for the genotype-environmental covariation which might be expected if cultural transmission were superimposed upon genetically determined differences (Eaves, 1976b).

2.2 Analysis of Variance

Simple one way analyses of variance between and within pairs of twins were carried out for each of the five twin types and the

TABLE B 21 : SIMPLE ONE-WAY ANALYSIS OF VARIANCE OF TWINS RADICALISM SCORES

GROUP	ITEM	DF	SUM OF SQUARES	MEAN SQUARE	F	P	r (intra-class correlation)
MZ MALES	between pairs	81	93.2946	1.1377	3.79	***	0.5824
	within pairs	83	24.9174	0.3002			
MZ FEMALES	between pairs	231	261.9857	1.1292	4.53	***	0.6381
	within pairs	233	58.1424	0.2495			
DZ MALES	between pairs	50	63.7308	1.2496	2.78	**	0.4714
	within pairs	52	23.3449	0.4489			
DZ FEMALES	between pairs	145	177.8865	1.2184	3.11	***	0.5134
	within pairs	147	57.5830	0.3917			
DZ OPPOSITE-SEX	between pairs	70	190.9946	1.2816	2.61	***	0.4411
	within pairs	72	35.3557	0.4911			

TABLE B 22 : SIMPLE ONE WAY ANALYSIS OF VARIANCE OF TOUGHMINDEDNESS SCORES

GROUP	ITEM	DF	SUM OF SQUARES	MEAN SQUARE	F	P	r (intraclass correlation)
MZ MALES	between pairs	81	10.1931	0.1243	5.69	***	0.7004
	within pairs	83	1.8148	0.0219			
MZ FEMALES	between pairs	231	28.3688	0.1224	3.57	***	0.5622
	within pairs	233	7.9998	0.0343			
DZ MALES	between pairs	50	6.3842	0.1252	3.46	***	0.5514
	within pairs	52	1.8826	0.0362			
DZ FEMALES	between pairs	145	13.8700	0.0950	2.43	***	0.4169
	within pairs	147	5.7522	0.0391			
DZ OPPOSITE-SEX	between pairs	70	7.8495	0.1106	2.78	***	0.4698
	within pairs	72	2.8682	0.0398			

results are presented in Tables B21 and B22. One degree of freedom has been lost between pairs in each case by the age-correction, which removed variation between pairs due to the regression of test score on age.

Variation was significantly greater between than within pairs in all groups for both traits, showing that variation is not the result of random differences between individuals. There must be environmental or genetical differences between families. Genetical variation may be indicated since the within pairs mean squares of identical twins are markedly less than those of non-identical twins.

The male within pairs variance is non-significantly larger than the female within pairs variance for both MZ and DZ twins. Conversely, variation in Toughmindedness is greater within female pairs. Evidence for a sex difference in the mechanisms determining individual differences in Radicalism and Toughmindedness is provided by the opposite-sex non-identical twins, whose within pairs variance is greater than that in either male or female DZ pairs for both traits.

Since the total variances of the sub-groups of the data are homogeneous, no information is lost in calculating correlation coefficients from the data. Before proceeding to a more formal investigation of the causes of individual differences in Radicalism and Toughmindedness, the correlation coefficients from the twin and adoption study will be considered, in order to determine what information they yield about individual differences in these data.

3 CORRELATIONS

3.1 Correlations of the Twins

The intraclass correlation coefficients, calculated from the variance components, are given in Tables B21 and B22. Small, but consistent differences are observed among the correlations. For both Radicalism and Toughmindedness, the MZ twins correlations are higher than those of the DZ twins suggesting either that there is greater genetical similarity between members of an MZ pair or that MZ pairs are treated more alike. A sex effect is observed for both traits. Correlations are higher in females than males for Radicalism. The reverse sex effect is observed for Toughmindedness.

We wish to quantify these observed differences in the correlations and determine whether they are statistically significant. Because the distribution of the correlation coefficient is skewed, the correlations were transformed into z values, which are normally distributed, and the z values were tested for homogeneity as described in Section 4 of Part A.

The five intraclass correlations differed significantly at the 5% level for Toughmindedness ($\chi^2_4 = 10.13$), but not for Radicalism ($\chi^2_4 = 6.57$). The differences between male and female correlations within zygosities noted earlier were not significant for either trait. Therefore, since there was no evidence for the effect of sex, the correlations for each zygosity were pooled over sexes. Pooled MZ and DZ correlations differed significantly at the 5% level for both traits.

Comparison of the pooled same-sex DZ twin correlation with the opposite-sex DZ twin correlation, showed no significant difference for either Radicalism or Toughmindedness. Thus, the correlations provide no evidence for a sex difference in the determination of attitudes, which had been suggested earlier.

A model specifying comparisons between zygosities, sexes and between same and opposite sex twins was written and is given in Table B23. The model was fitted to the z values for Radicalism and Toughmindedness using weighted least squares, as described in Section 4 of Part A. The results are presented in Table B23. The only significant effect for both traits is the mean effect. The overall effects of zygosity, sex and the same-sex vs. opposite-sex comparison were not significant. However, the relative magnitudes of these parameters is informative. Apart from the mean, the zygosities effect is the most substantial effect for both traits. For Toughmindedness, the sex effect is of a similar magnitude. However, the sex effect for Radicalism and the same-sex vs. opposite-sex comparison for both traits are negligible.

We have evidence for significant differences between the MZ and DZ twin correlations from the test of homogeneity of the MZ and DZ twin correlations pooled over sexes. This indicates that there may be a genetical component of variation for these traits, despite the lack of an overall zygosities effect in this analysis.

TABLE B 23 : MODEL AND ESTIMATES FOR CORRELATIONS OF TWINS

GROUP	MEAN	ZYGOSITY	SEX	SAME-SEX VS. OPPOSITE-SEX
MZ Males	1	1	1	0
MZ Females	1	1	-1	0
DZ Males	1	-2/3	1	1
DZ Females	1	-2/3	-1	1
DZ opposite-sex	1	-2/3	0	-2
RADICALISM				
Estimate	0.5945	0.1178	-0.0377	0.0212
Standard Error	0.1441	0.1674	0.1576	0.1386
Significance level	***	NS	NS	NS
TOUGHMINDEDNESS				
Estimate	0.6160	0.1329	0.1038	0.0088
Standard Error	0.1441	0.1674	0.1576	0.1386
Significance level	***	NS	NS	NS

3.2 Correlation among the individuals of the Adoption Study

Pearson correlation coefficients were calculated among the individuals of the Adoption Study. The 6 x 6 correlation matrices for Radicalism and Toughmindedness are given in Tables B25 and B24.

There are significant correlations between spouses of 0.4045 and 0.3824 for Radicalism and Toughmindedness respectively, as would have been predicted from previous work. However, these correlations are somewhat lower than the values of about 0.6 reported by Insel (1974) and others, but do not differ significantly from them.

The only other significant correlation for Radicalism is that between adopted daughter and adopting mother, which is based on the largest number of pairs of observations. This correlation suggests that it will be interesting to consider the pattern of the other correlations because although formally they do not differ from zero, they are not significantly different from the foster mother-adopted daughter correlation either.

Firstly, considering Radicalism, the correlations of adopted sons and daughters with their adopting mothers are very similar, whereas those with their fathers differ somewhat (and are not significantly greater than zero). Pooling over sexes, the correlation of an adopted child with its mother is higher than that with its father, suggesting that the role of a mother in providing a child's environment may be more significant in the determination of Radicalism, than the role of a father. Correlations between adopted and natural children reared by the same parents are slightly higher than those between fosterparent

TABLE B 24 : CORRELATION AMONG RADICALISM SCORES

Adopted Male	Adopted Female	Adopting Mother	Adopting Father	Natural Son of Adopting Parents	Natural Daughter of Adopting Parents	
1.0000	- (0)	0.3160 (9)	0.5153 (5)	- (0)	- (1)	Adopted Male
	1.0000	0.3363* (40)	0.0729 (34)	0.3579 (8)	0.4480 (8)	Adopted Female
		1.0000	0.4045** (47)	0.7554 (3)	0.3679 (5)	Adopting Mother
			1.0000	- (2)	0.5057 (5)	Adopting Father
				1.0000	- (1)	Natural Son of Adopting Parent
					1.0000	Natural Daughter of Adopting Parent

NOTES :

1. Numbers in brackets give number of pairs on which the correlation was based.

2. Key

   * Significant at the 5% level

   ** Significant at the 1% level

   - No correlation could be calculated

TABLE B 25 : CORRELATION AMONG TOUGH-MINDEDNESS SCORES

Adopted Male	Adopted Female	Adopting Mother	Adopting Father	Natural Son of Adopting Parents	Natural daughter of adopting parents	
1.0000	-	0.3369 (9)	0.6468 (5)	-	-	Adopted Male
	1.0000	0.0095 (40)	0.2294 (34)	0.5723 (8)	0.1787 (8)	Adopted Female
		1.0000	0.3824** (47)	0.8582 (3)	0.5243 (5)	Adopting Mother
			1.0000	- (2)	0.6081 (5)	Adopting Father
				1.0000	- (1)	Natural son of adopting parents
					1.0000	Natural daughter of adopting parents

and adopted child. Correlations between parents and their natural children are higher than those between adopting parents and their adopted children, which although based on small numbers is consistent with the hypothesis that some of the similarity between natural parents and offspring may be genetic in origin.

Turning now to Toughmindedness, only the marital correlation is significant. The correlations of natural children with their parents tend to be higher than the others, again suggesting a genetical component, but this is all that we can usefully say about them.

The inter-correlations among Radicalism, Toughmindedness and Age, presented in Table B26, are mostly non-significant. However, a few comments will be made. Not surprisingly there is a significant correlation in the age of spouses and also between the ages of parents and adopted children. Two significant correlations between subjects' Radicalism and Toughminded scores are observed, which reflect the non-orthogonality of the traits ($r = 0.19$). Since the data have been age-corrected, we do not expect to find correlations between Radicalism or Toughmindedness and age. However, the correlation between age of mother and Toughmindedness in their adopted daughters is negative and significant at the 5% level. Older mothers who are more Tenderminded themselves, may provide an environment in which their offspring tend to be more tenderminded. However, since we expect one correlation in twenty to be significant at the 5% level, we cannot attach too much importance to this finding.

Overall, looking at the correlations has not been very helpful, apart from demonstrating a marital correlation, because the numbers on which they are based are too few. The causes of variation in the Adoptian Study may be elucidated by using all the information available simultaneously, which we shall do in a later section.

TABLE B 26 : CORRELATIONS BETWEEN RADICALISM, TOUGHMINDEDNESS AND AGE

		TOUGHMINDEDNESS						AGE					
		1	2	3	4	5	6	1	2	3	4	5	6
Age	1	-06	-	03	69	-	-	100	-	70 *	52	-	-
	2	-	01	-04	-24	-47	-37	-	100	33 *	61 **	81 *	27
	3	-41	-34 *	-06	-04	-83	-15	70 *	33 *	100	77 **	-25	71
	4	-32	-20	-17	-21	-	-30	52	61 **	77 **	100	-	14
	5	-	-05	31	-	-10	-	-	81 *	-25	-	100	-
	6	-	-17	-39	-07	-	-30	-	27	71	14	-	100
Radicalism	1	-01	-	02	69	-	-	03	-	27	-28	-	-
	2	-	21 **	03	-03	04	11	-	02	-09	17	12	-08
	3	20	14	06	16	85	76	04	-17	-13	-03	30	-68
	4	47	07	14	19	-	-02	46	-00	13	13	-	-78
	5	-	70 +	76	-	82 **	-	-	-00	-73	-	36	-
	6	-	26	01	01	-	20	-	58	-02	15	-	-36

NOTES :

1. All decimal points are omitted

2. Code numbers are :
   1 Adopted Male
   2 Adopted Female
   3 Adopting Mother
   4 Adopting Father
   5 Natural son of adopting parents
   6 Natural daughter of adopting parents

SECTION 6: SIMPLE MODELS OF VARIATION IN RADICALISM AND TOUGHMINDEDNESS IN THE TWINS

1 MODEL FITTING PROCEDURES

Throughout this Section, the concepts and methods discussed at length in Section 3 of Part A will be employed to investigate the genetical and environmental mechanisms, producing the pattern of mean squares observed in the twins. This will enable us to compare this study with previous studies, before moving on to examine all the data simultaneously, using methods appropriate to the analysis of unbalanced pedigrees.

Expectations for the ten mean squares will be generated in terms of the genetical and environmental components of variation described in Part A. The most likely explanation of the observed pattern of variation is that employing the least number of parameters representing sources of variation, in a model which satisfies the chisquare criterion of adequacy. Therefore, we first test the adequacy of the simplest models of variation and only incorporate additional parameters into our models when the simpler models are inadequate to explain variation in the data, or when the fitting of an additional parameter produces a significant improvement in fit over simpler models. Thus, certain effects influencing a trait may not be detected and estimated if they do not account for a significant proportion of the variation. The contribution of such effects to the total variation will be small and our explanation of the variation within a parsimonious model will account for the major part of the variation. Biases introduced by accepting as adequate models which are too simple will be calculated in Part C,

where the sample sizes needed to detect different components of variation will also be given.

2 MODELS OF RADICALISM IN THE TWINS

The simplest model of variation, the E_1 model, specifies that all variation is due to random effects such as errors of measurement or environmental influences specific to the individual. This model clearly fails to explain the observed variation ($\chi^2_9 = 178.03$) and its failure indicates that $\sigma^2_b \neq \sigma^2_w$ and $\sigma^2_{wMZ} = \sigma^2_{w\,DZ}$. Environmental differences, such as cultural differences, between families (E_2) might produce the former inequality and differences in E_1 between MZ and DZ twins the latter. Genetical variation would produce both inequalities.

In order to determine whether either genetical or environmental differences between families alone is sufficient to account for the observed residual variation after fitting E_1, two simple two-parameter models were tested. The simple environmental model (E_1B) specifies that all variation may be accounted for by environmental differences either within or between families. We see from Table B27, that this model is not adequate since significant residual variation remains after fitting it. The simple genetical model (E_1D_R) specifies specific environmental effects, and errors of measurement and also additive genetical variation. Other sources of environmental variation (e.g. cultural differences) and genetical variation (i.e. all non-additive effects: dominance, epistasis genotype-environment interaction, genotype-environment covariation) will lead to failure of this simple model. From Table B27, we see that this model yields a non-significant chisquare. However, the fit is poor indicating that

TABLE B 27 : TWO SIMPLE MODELS FOR RADICALISM IN THE TWINS

1. SIMPLE ENVIRONMENTAL MODEL

Parameter	Estimate	Standard Error	Significant Level
E_1	0.3396	0.0198	$p < 0.001$
B	0.4211	0.0361	$p < 0.001$

$\chi^2_8 = 21.029, \quad p = 0.007$

2. SIMPLE GENETICAL MODEL

Parameter	Estimate	Standard Error	Significant Level
E_1	0.2536	0.0193	$p < 0.001$
D_R	1.0068	0.0764	$p < 0.001$

$\chi^2_8 = 14.276, \quad p = 0.075$

Narrow Heritability $(h^2_N) = 0.6650 \pm 0.0284$

there might be an additional source of variation.

The most obvious three parameter model to try is that incorporating all three sources of variation described so far i.e. the E_1D_RB model. However, when there is genetical variation, interpretation of the B parameter changes. This parameter specifies between families variation, which may be produced by cultural differences between families as specified in the E_1B model. A priori, we might expect cultural differences to produce variation in Radicalism. However, since the E_1D_RB model includes additive genetical variation, genetical variation produced by such factors as dominance may contribute to the between families component. Eaves (1972) has shown that the power of the test for detecting dominance, using twins reared together, is extremely low and we would be unlikely to pick up its effect. However, in this study another effect must be considered.

We know from the adoption study that there is a correlation of 0.4045 between spouses in their Radicalism scores. Cultural effects and assortative mating both contribute to variation between families of individuals reared together. If cultural differences are perpetuated by cultural transmission (Eaves, 1976b), then the evolutionary consequences of cultural transmission and assortative mating are likely to be similar in the presence of genetical variation, since both increase apparent genetical variation. Cultural transmission does this by perpetuating environmentally the consequences of genetical segregation in previous generations. Assortative mating does so by associating alleles of like effect. So, when we are looking at twins reared together, we can only determine whether the effects of the mating system and culture are playing a jointly significant role in the determination of differences between families.

The joint contribution of assortative mating and cultural differences is thus summarised by the parameter B in the E_1D_RB model, which is given in Table B28.

The results of fitting this model to the data are given in Table B29. This model provides an adequate explanation of the data ($\chi^2_7 = 6.198$) and is a significant improvement over the E_1D_R model ($\chi^2_1 = 8.078$). Despite the fact that we have included all three parameters in this model, each differs significantly from zero and, therefore, represents a significant source of variation. The data are consistent with the interpretation that specific environmental influences and additive genetical effects produce variation in Radicalism. These findings are of limited interest since both D_R and E_1 are random, non-directional effects which could produce variation in a trait that is not biologically or sociologically important (although it is possible that such a trait could be under stabilising selection). However, if there is assortative mating, it seems likely that the trait may be of some evolutionary consequence.

Both assortative mating and cultural differences show that the trait may be sociologically relevant.

We know that either assortative mating or cultural effects or both are producing variation between families, but we cannot determine their relative contribution using the twin data alone. We can only set upper and lower limits for their effects. The total variance is given by $(\frac{1}{2} D_R + E_1 + B)$. Therefore, the parameters account for the following percentage of the total variation:

TABLE B 28 : MODEL FITTED TO MEAN SQUARES FOR RADICALISM

TWIN TYPE	MEAN SQUARE	EXPECTED MEAN SQUARE		
		D_R	E_1	B
MZ_m	b	1	1	2
	w	-	1	-
MZ_f	b	1	1	2
	w	-	1	-
DZ_m	b	$\frac{3}{4}$	1	2
	w	$\frac{1}{4}$	1	-
DZ_f	b	$\frac{3}{4}$	1	2
	w	$\frac{1}{4}$	1	-
DZ_{mf}	b	$\frac{3}{4}$	1	2
	w	$\frac{1}{4}$	1	-

where b = between pairs mean square

w = within pairs mean square

TABLE B 29 : FINAL MODEL FOR RADICALISM IN THE TWINS

Parameter	Estimate	Standard Error	Significance Level
E_1	0.2687	0.0212	$p < 0.001$
D_R	0.5765	0.1589	$p < 0.001$
B	0.2091	0.0734	$p < 0.01$

$\chi^2_7 = 6.198$, $p = 0.517$

$$\tfrac{1}{2}D_R = 37.63\%$$
$$E_1 = 35.08\%$$
$$B = 27.30\%$$

If the contribution to B of assortative mating is zero, then cultural differences may account for as much as 27.30% of the total variation and 37.63% of the variation in Radicalism could be due to genetical differences. However, we know from the adoption study that there is assortative mating. If we now assume assortative mating such that B is entirely genetic in origin, then as much as 64.92% of the variation could be due to genetical differences. In other words anything between 38 and 65% of the variation could be due to genetical differences.

Since we have data on spouses, we can go further than this and resolve the relative contributions of assortative mating and cultural effects. We will take the value of the marital correlation, μ, as 0.4045 and assume that the only influence of parents on their children is genetic. Assortative mating increases the additive genetical variance by associating alleles of like effect. Fisher showed that under certain conditions the contribution of assortative mating to the additive genetical variance is given by $\frac{1}{2}(A/(1-A))D_R$, where A is the correlation between the additive genetical deviations of spouses. "A" can now be calculated from μ and the narrow heritability (h^2_N), where h^2_N is the proportion of the total variation due to additive effects and is given by:

$$h^2_N = \frac{\frac{1}{2}D_R + \frac{1}{2}(A/(1-A))D_R}{V_T}$$

where $V_T = \frac{1}{2}D_R + E_1 + B$

This simplifies to:

$$h^2_N = \frac{\frac{1}{2} D_R((1/(1 - A))}{V_T}$$

Making the usual assumption that assortative mating is based on the phenotype, then:

$$A = \mu h^2_N$$

We have estimates of μ and $\frac{1}{2} D_R/V_T$ and can use these to obtain a quadratic in A:

$$A = 0.4045 \times 0.3763 \times 1/(1 - A)$$

which simplifies to:

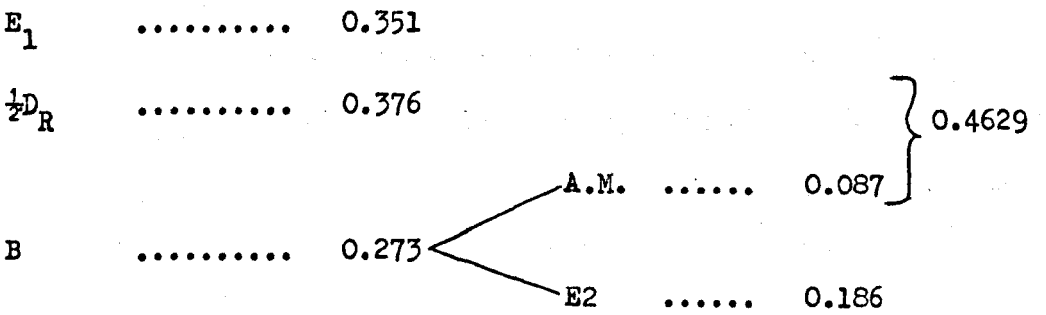
$$A^2 - A + 0.1522 = 0$$

Solutions for A are 0.8127 and 0.1873. Since the first value is greater than μ, we take A = 0.1873. Therefore, assortative mating accounts for $\frac{1}{2} D_R(A/(1 - A))/V_T = 8.67\%$ of the total variance. The difference between this value and the percentage of V_T accounted for by B is 18.63%, which is variation due to environmental differences between families. We can determine the proportion of the total variation due to genetical differences by adding the contribution of assortative mating to that which would still persist if mating were random. Thus, genetical differences account for 46.29% of the total variation.

A summary of our final model for the mechanism producing variation in Radicalism is summarised in Table B30. We now wish to compare these results with those of previous studies and then to make

TABLE B 30 : FINAL ANALYSIS OF RADICALISM SCORES FOR 587 PAIRS OF TWINS

Estimates of the components as a percentage of the total variation:



N.B. A.M. is the percentage of the total variance due to assortative mating.

i.e. A.M. $= \frac{1}{2}D_R \ (A/(1-A)) \ /V_T$

predictions about the pattern of variances and covariances expected in the adoption study. If the adoption data are consistent with these predictions, we may obtain estimates of genetical and environmental components of variation using all the data simultaneously.

In summary, the most significant feature of this analysis of Radicalism is that the simple models fail. For many traits the simple models fit (see Martin,1977 for many examples). In some cases, lack of ability to discriminate between simple environmental and simple genetical models indicates that the data are inadequate. However, in others the adequacy of the simple models suggests either that many of the complex mechanisms hypothesised for human variation do not operate in reality or that the traits being studied are not very relevant for our understanding of biology or society.

We have shown that the genetical effects of the mating system and cultural or treatment differences between families are contributing significantly to variation in Radicalism. These effects could reflect the cultural impact of parents on their children. Thus, Radicalism may be a sociologically relevant trait of evolutionary significance, where the joint role of cultural differences and heredity in the transmission of culture can be studied.

3 COMPONENTS OF VARIATION IN RADICALISM IN THREE STUDIES

The results for Radicalism of three studies of social attitudes are presented in Table B31, where the mean squares and proportion of the variation accounted for by the three parameters required to account for the variation are given. These studies (Hewitt, 1974; Martin, 1976; and the present study) are surprisingly consistent, despite the fact that the data were collected at different times, using different questionnaires and with very different aims in mind. Hewitt (1974) states in his thesis that he wished to determine whether the structure of attitudes had altered since the earlier work summarised by Eysenck (1954), as had been suggested by Eaves and Eysenck (1974a), and to examine subjects x items interactions for attitudes using the approach of Eaves and Eysenck (1974b). The collection of this data was initiated by Eaves who wished to study the relationship between personality and social attitudes. Martin's (1976) data were obtained as part of a study designed primarily to examine the determination of individual differences in sexual attitudes and the relationship of these attitudes with other behavioural traits. The data of this twin study were obtained in conjunction with data on smoking and drinking behaviour, neuroticism and impulsiveness. As a result of the consistency of a preliminary analysis of this study by Eaves (personal communication) with that of the other twins studies, the data of the Adoption Study were obtained in order to test hypotheses formulated on the basis of the twin data and to resolve parameters which are inevitably confounded in data on individuals reared together.

TABLE B 31 : COMPARISON OF THREE STUDIES OF RADICALISM

BETWEEN PAIRS MEAN SQUARES		HEWITT df	HEWITT MS
MONOZYGOTIC	MALES	141	10.3111
	FEMALES	323	8.36411
DIZYGOTIC	MALES	36	8.8544
	FEMALES	193	9.5146
	OPPOSITE-SEX	126	10.1125
WITHIN PAIRS MEAN SQUARES			
MONOZYGOTIC	MALES	142	1.7804
	FEMALES	324	1.9055
DIZYGOTIC	MALES	37	3.2145
	FEMALES	194	2.8950
	OPPOSITE-SEX	126	3.2892

% of Total Variation Accounted For By :

E_1	0.3329	
$\frac{1}{2}D_R$	0.3541	
B	0.3130	AM—0.156, E_2—0.157
μ	0.60	

MARTIN		PRESENT STUDY	
df	MS	df	MS
37	356.521	81	1.1377
93	337.97	231	1.1292
15	272.48	50	1.2496
52	365.26	145	1.2184
39	350.61	70	1.2816
39	49.20	83	0.3002
95	61.50	233	0.2495
17	82.38	52	0.4489
54	100.83	147	0.3917
41	129.35	72	0.4911
0.2736		0.3508	
0.4427		0.3763	
0.2736	AM – 0.2736 E_2 – 0.0000	0.2730	AM — 0.087 E_2 — 0.186
0.67		0.4045	

Examination of the between pairs mean squares in Table B31 reveals no similarity in their relative magnitudes in the five twin groups in the three studies. However, in all cases, these mean squares were homogeneous and, therefore, the observed patterns in the three studies can be attributed to chance factors. Within pairs mean squares conform to the pattern we expect if there are genetical differences producing variation in Radicalism. MZ twin mean squares are substantially less than DZ twin mean squares. The mean squares within opposite-sex DZ pairs are slightly larger than those within same-sex DZ pairs, indicating slight but non-significant heterogeneity between the sexes.

In all three studies, the most appropriate explanation of the observed pattern of mean squares was found in a model incorporating specific environmental, cultural (or assortative mating) and additive genetical differences. These parameters accounted for a substantial proportion of the total variation in each case.

We first compare the results of the two largest and most reliable studies - Hewitts, based on 824 pairs of twins and the present study based on 587 pairs. These only suffer from the usual problems of sampling associated with twin studies, whereas the response rate in Martin's study was poor (only 30%) because the social attitudes questionnaire was sent out with a questionnaire on sexual attitudes. Problems will be introduced by the low response rate and the likely biases in the type of individual who would return such an anonymous questionnaire.

Hewitt's and the present study yield almost identical estimates of $\frac{1}{2}D_R$ when these are expressed as proportion of the total variance. $\hat{E}_1$ is slightly larger and $\hat{B}$ slightly smaller in the present study. The

authors' estimates of the relative contributions of assortative mating and cultural effects differ somewhat, but this is explained by the estimates they choose for the value of the marital correlation. Hewitt employs a value suggested by Insel's (1961) study, using the Wilson Conservatism scale, of $\mu = 0.60$. This yields a solution where cultural effects and assortative mating make an equal contribution to the variation. The present analysis uses a lower value of $\mu = 0.40$, obtained from the adoption study, and it is found that cultural effects contribute more than twice as much as assortative mating to the total between families variation. If a value of $\mu = 0.40$ is used with Hewitt's data, rather than the value of $\mu = 0.60$ which he chooses, then cultural effects account for 23.9% of the variation and assortative mating for 7.4% of the total variation, a result very similar to our own. Thus, as is expected, the precise contributions of assortative mating and cultural effects, estimated from the data, depends on the value of μ used. We believe that Insel's correlation may be rather high (as are other correlations he reports; see Eaves and Eysenck (1974)) and that cultural effects may produce more variation in Radicalism than assortative mating. If we assume that $\hat{B}$ is all variation produced by assortative mating, we can easily calculate the value of μ that would be found in the population. For Hewitt's study $\mu = 0.70$ and for the present study $\mu = 0.65$. This value in Hewitt's study is higher than any reported before and the value in our study is considerably higher than the observed correlation. Therefore, although the exact size of their relative contributions is still in question, both cultural effects and assortative mating seem to be playing a role in determining individual differences in Radicalism.

Martin's study might, a priori, be expected to be less reliable because of the sampling procedure and the smaller sample size (246 pairs). Here the value of $\hat{B}$ is about the same as in the other two studies, but $\frac{1}{2}\hat{D}_R$ is larger and $\hat{E}_1$ is smaller. These estimates are not significantly different to those from the other studies, and overall the studies are in good agreement with one another. However, the size of $\frac{1}{2}\hat{D}_R$ has interesting implications when we look at the distribution of $\hat{B}$ between cultural effects and assortative mating. Martin uses a value of $\mu = 0.67$, reported by Eysenck (1976) for the 88 item Public Opinion Inventory used in his study, in order to calculate the proportion of $\hat{B}$ which can be accounted for by assortative mating. He finds that the quadratic in A has no real solution since $\mu \times \frac{1}{2} D_R/V_T \times 4 > 1$ and concludes that his data are compatible with the explanation that all the between families variation can be accounted for by assortative mating. Assuming again that there is no variation produced by cultural effects and that all $\hat{B}$ is due to assortative mating, we estimate the value of μ as 0.53, which is lower than that used by Martin and explains why no real solution existed for the quadratic. This sets a maximum value of $\mu = 0.53$ for Martin's study (i.e. given that Fishers model is appropriate) and once again suggests that Insel's reported correlations are unusually high or that an entirely different model need be adopted for the findings (e.g. genotype-environment covariance such as that produced by cultural transmission) Using $\mu = 0.4045$ as for the present study, Martin's data yield estimates of assortative mating = 13.5% and cultural effects = 13.9% of the total variation. Thus, depending on which value of μ we use in

the three studies we can get almost any relative contribution of assortative mating and cultural effects to the variation between families. However, from the two largest studies, we believe that both are presented in some degree.

The overall consistency of the three studies, despite the various inadequacies reported by their authors, suggests that our final model of Radicalism in some measure approximates to the true answer and gives some confidence to our efforts to generalise from the study of twins. It now remains to test the generality of this model for other types of individual in order to decide whether or not twins are typical of the population as a whole with respect to the genetical and environmental influences operating on them. We note in passing that so far we have found no evidence for one influence which could differentiate twins from other types of family for this trait, namely competition or co-operation, since the total variances of all twin groups were homogeneous. The adoption data will be used to clarify the position with respect to cultural effects and assortative mating and to test the applicability of the model in the twins to the population as a whole.

4 MODELS OF TOUGHMINDEDNESS IN THE TWINS

Using the same approach as that adopted for Radicalism, the simple E_1 model was first fitted to the data and failed to account for the observed variation. The results of then fitting the two parameter models are given in Table B32. The simple genetical model does not provide an adequate explanation of the data. However the simple environmental model appears to fit quite well, with 46.24% of the variation accounted for by specific environmental factors and 53.40% by between families differences. Therefore, formally at least, we need seek no further than these simple environmental effects for causes of variation in Toughmindedness.

However, as for the previous trait, we fitted the three parameters, E_1, B and D_R to the data. Although all three parameters were significant, the additional parameter just failed to produce any significant improvement in fit over the simple environmental model ($\chi^2_1 = 3.71$, $5\% < p < 10\%$).

Thus, formally the E_1E_2 model is adequate to account for variation in the twins. However, addition of the D_R parameter produces a substantial although not significant reduction in chisquare and this parameter is significantly greater than zero. Therefore, a small genetical component of variation may be indicated, which might be detected in a larger study of this type or a study with a completely different design (e.g. the Adoption Study).

How do these results compare with those of the two previous twin studies described earlier? The means squares for the three studies are presented together in Table B33 for comparison. Hewitt (1974) found

TABLE B 32 : TWO SIMPLE MODELS FOR TOUGHMINDEDNESS IN THE TWINS

1. Simple Environmental Model

Parameter	Estimate	Standard Error	Significance Level
E_1	0.03461	0.00202	$p < 0.001$
E_2	0.03997	0.00352	$p < 0.001$

$\chi^2_8 = 10.999$, $p = 0.202$

2. Simple Genetical Model

Parameter	Estimate	Standard Error	Significance Level
E_1	0.02875	0.00216	$p < 0.001$
D_R	0.08902	0.00744	$p < 0.001$

$\chi^2_8 = 18.224$, $p = 0.020$

Narrow Heritability $(h^2_N) = 0.6076 \pm 0.0322$

TABLE B 33 : COMPARISON OF THREE STUDIES OF TOUGHMINDEDNESS

BETWEEN PAIRS MEAN SQUARES		HEWITT	
		df	MS
MONOZYGOTIC	MALES	141	7.8197
	FEMALES	323	10.3697
DIZYGOTIC	MALES	36	6.5694
	FEMALES	193	7.5213
	OPPOSITE-SEX	126	8.1251
WITHIN PAIRS MEAN SQUARES			
MONOZYGOTIC	MALES	142	3.1980
	FEMALES	324	1.8643
DIZYGOTIC	MALES	37	3.3720
	FEMALES	194	2.9982
	OPPOSITE-SEX	126	4.3155

MARTIN		PRESENT STUDY	
df	MS	df	MS
37	391.03	81	0.12431
93	383.21	231	0.12241
15	548.82	50	0.12518
52	248.89	145	0.09500
39	299.19	70	0.11056
39	93.36	83	0.02187
95	79.48	233	0.03433
17	62.91	52	0.03260
54	125.14	147	0.03913
41	148.63	72	0.03984

that no model with the same effects in both sexes would fit these data. The simple genetical model was adequate to explain the variation in each sex separately, but the simple environmental model again failed. Eaves (personal communication) later devised a model which could be fitted to both sexes simultaneously by allowing the parameters to take different values in each sex and including a parameter to specify the correlation between an effect in males and females. He found, for Hewitt's study, that a model with separate E_1's (E_{1M} and E_{1F}), separate D_R's (D_{RM} and D_{RF}) and an interaction parameter (D_{RMF}) was adequate to explain the variation, but that the fit was poor ($\chi^2_5 = 9.51$: $0.05 < p < 0.10$) (Eaves, 1977). Addition of a parameter summarising cultural differences (or assortative mating) - B - gave a significant improvement in fit over the earlier model and suggested that the mechanism producing variation in the two sexes was very different. Eaves concluded that the principal characteristics of Toughmindedness were those of a trait whose pattern of determination shows a marked interaction with sex. Both sexes seemed to be subject to the effects of the family environment (since $\hat{B}$ was significant). However, there was only evidence for genetical variation in females. The genes did not appear to be contributing to individual differences in males.

Martin's (1976) study yielded very similar results. Different models of variation were required in males and females. He found that of the two simple two parameter models only the E_1E_2 model would fit in males and only the E_1D_R model would fit in females, which was consistent with the findings of Eaves for Hewitt's data. He found that a similar model to that of Eaves (allowing the parameters to take separate values in the two sexes and allowing for interaction between an effect in males and females) gave the best explanation of his data.

Thus, in both previous studies no simple model would account for variation in both sexes, and the determination of attitudes was markedly different in the two sexes. There was a heritable component of variation in females, but no evidence for genetical variation in males. In this study, there was no heterogeneity between the sexes and no significant genetical component of variation was found in either sex. How can we explain the discrepancy between the two previous studies, which agreed with one another well, and the present study?

The most obvious culprit in such cases is biased sampling. However, this does not seem to be an appropriate reason here, since the findings in all three studies for Radicalism were remarkably consistent. For our explanation, we must turn, I think, to the questionnaire used, since this is the most obvious difference between the previous studies and the present study. Such problems as biased sampling were common to all three studies.

Hewitt's and Martin's studies used different questionnaires, but both were versions of the Public Opinion Inventory where subjects were asked to rate 80 statements from 1 (strongly disagree) to 5 (strongly agree). The present study used the Wilson-Patterson attitudes questionnaire where subjects respond by agreeing or disagreeing with a topic represented by a single noun. This questionnaire was developed from the Wilson Conservatism Scale (1968). A comparison of the Wilson-Conservatism scale and the Wilson-Patterson Attitudes Questionnaire is given for interest in Table B34. Wilson and Patterson (1968) state about the origin of the Conservatism scale that: "Development of the present scale began with a list of characteristics which might be expected in the extreme conservative". Fifty items which were effective

TABLE B 34 : COMPARISON OF WILSON'S (1968) CONSERVATISM SCALE AND THE WILSON-PATTERSON SOCIAL ATTITUDES QUESTIONNAIRE

ITEMS SPECIFIC TO EARLIER CONSERVATISM SCALE	ITEMS COMMON TO BOTH SCALES	ITEMS SPECIFIC TO SOCIAL ATTITUDES SCALE
Chaperones	1. Death Penalty	5. Council Housing
Student Pranks	2. Evolution Theory	8. Women's Lib
Fluoridisation	3. School Uniforms	9. Student Protest
White Lies	4. Striptease Shows	10. Police
Straitjackets	6. Sunday Observance	11. Surtax
Learning Latin	7. Hippies (Beatniks)	13. Foreign Aid
	12. Wife Swapping (Pyjama Parties)	15. Welfare State
	14. Pop Music (Jazz)	16. Tradition
	18. The Pill	17. Conscription
	19. Patriotism	21. United Nations
	20. Modern Art	26. Law Reform
	22. Self Denial	33. Political Systems
	23. Working Mothers	37. Space Research
	24. Military Drill	38. Strikes
	25. Co-Education	39. Common Market
	27. Divine Law	44. Capitalism
	28. Socialism	50. Inheritance Tax
	29. White Superiority	57. Arab Politics
	30. Cousin Marriage	58. Pacificism
	31. Moral Training	59. Law and Order
	32. Suicide	62. Profit Sharing
	34. Legalised Abortion	66. Trade Unions
	35. Empire Building	67. Supersonic Airliners
	36. Licensing Laws	68. Liberals
	40. Computer Music	
	41. Chastity	
	42. Royalty	
	43. Women Judges	
	45. Conventional Clothing	
	46. Teenage Drivers	
	47. Apartheid	
	48. Nudist Camps	

contd.

TABLE B 34 Contd.

ITEMS SPECIFIC TO EARLIER CONSERVATISM SCALE	ITEMS COMMON TO BOTH SCALES	ITEMS SPECIFIC TO SOCIAL ATTITUDES SCALE
	49. Church Authority	
	51. Astrology (Horoscopes)	
	52. Disarmament	
	53. Censorship	
	54. Birching	
	55. Mixed Marriage	
	56. Strict Rules	
	60. Casual Living	
	61. Divorce	
	63. Inborn Conscience	
	64. Coloured Immigration	
	65. Bible Truth	

discriminators on these characteristics, were chosen from a large pool of items, so that half the items scored positively on Conservatism and half negatively. Examination of the items added to this conservatism scale to produce the questionnaire used in the present study does not suggest that they were especially chosen to discriminate between Tough and Tendermindedness. This may be the key to the difference between the present study and earlier studies. Hewitt (1974) reports from his factor analysis of the 60 item Public Opinion Inventory that the first principal component accounted for 14% of the total variation, the second for 10% and all remaining components for 3% or less. Other studies using similar questionnaires to the 60 item Public Opinion Inventory, find that Radicalism and Toughmindedness are two major factors approximately equal magnitude. Eysenck does not report the proportion of variation accounted for by R and T in his analysis of the quota sample described previously. Although the item content and the loadings in the two factors in the present study are closely similar to those found by Eysenck in the quota sample, we do not find that the two factors are of a similar magnitude. The first principal component accounts for 10.6% of the total variation, the second and third for 4.4% and 3.6% and all subsequent components for less than 2%, suggesting that either one or three factors might provide the best description of the items in this study. The picture that emerges for Radicalism, the major factor in this study, agrees remarkably well with that found for previous studies, suggesting that the trait being measured is very similar in all three studies and that a model including specific environmental effects, cultural effects, additive genetical variation and assortative mating may be a close approximation to the true causes of variation in this trait.

In the present study, "Toughmindedness" seems to be only a minor factor and it seems likely that we are not dealing with the same trait as in previous studies. This can be explained by the particular items used in the present scale which emphasize the discrimination between Radicalism and Conservatism. The major factors of Toughmindedness described by Hewitt and Martin show a marked interaction with sex in their determination. Our conclusion is that the trait measured in the present study, which resembles Toughmindedness, is only a minor component of variation. The causes of variation may be much simpler than those for Toughmindedness as measured by conventional types of questionnaires. There is little evidence for a heritable component of variation in this twin study, and the variation can be accounted for by unreliability of measurement, specific environmental influences and environmental differences between families. Although the trait is only a minor component of the attitudes assessed in the Wilson-Patterson Questionnaire, there are differences between families which require explanation. We shall examine in detail the between families component for both "Toughmindedness" and "Radicalism" by now turning to the Adoption Study.

SECTION 7: FITTING MODELS SIMULTANEOUSLY TO THE TWIN DATA AND THE UNBALANCED PEDIGREES OF THE ADOPTION STUDY

1 THE PROBLEM OF UNBALANCED PEDIGREES

The advantages of this study for the resolution of between families effects were discussed in Section 1. Analysis of the twin data has given a broad outline of the major causes of variation in attitudes. Significant between families variation was demonstrated, but we were unable to discriminate between certain hypotheses of considerable interest concerning this variation. For example, the relative contribution of variation produced by assortative mating and cultural effects could not be determined.

Cultural transmission is of great theoretical interest and, a priori, we might expect it to be of some importance in the determination of social attitudes. But since all between families components are inevitably confounded in studies of related individuals reared together, the twin study enabled us to show that cultural differences might contribute to the significant between families component but could provide no positive evidence for the presence of cultural transmission. Many writers have shown that data on adopted individuals provide the most direct test for between families environmental effects. The covariation of genetical and environmental effects which may be generated by the cultural transmission of inherited differences cannot be detected easily without adoption data (Jinks and Fulker, 1970). Cavalli-Sforza and Feldman (1973) comment on the importance of adopted individuals for detection of the components of their model of cultural transmission based

on a single gene. Therefore, in the present study, we hope to use the adoption data to estimate the parameters of a cultural transmission model for attitudes and compare this model with other simpler models in order to determine the likelihood that the non-genetic transfer of information from parents to offspring plays a significant role in the determination of attitudes.

Unfortunately, the design of our study is far from ideal. We have no data on the natural families of adopted children, and information on adoptive families is far from complete. We have a large number of adopted individuals, the adopting parents of a small proportion of these and a few natural and adopted siblings. All these individuals yield some information on effects of interest to us. Our problem is to make use of all this information.

Previously we have considered only balanced pedigrees. The structure of the twin data is consistent from family to family, with no missing observations. Such data are amenable to the analysis of variance. Analysis of such summary statistics as means, correlations and variances was presented in Sections 5 and 6. The procedures used were quick and convenient. We shall show later that using mean squares rather than working on the individual observations involves no serious loss of information when dealing with such regular pedigrees.

In Section 1 we saw that very few pedigrees of the adoption study had a similar structure. Such extremely unbalanced pedigrees present problems when attempting to use conventional summary statistics. Family sizes differ and the same individual may appear in a variety of relationships. Covariances and correlations between relatives are not independent and our usual approach is no longer appropriate.

Data on a large number of individuals with no measured relatives cannot contribute to correlations and covariances but contains a lot of information about total variances of different groups, which would be wasted using the conventional model fitting approach.

We shall present a method which enables the information provided by every individual to be used simultaneously, whatever his position in a particular pedigree. Single individuals are considered to be pedigrees of size one. Because of the awkward experimental "design" of this study, we shall abandon attempts to estimate conventional summary statistics and return to the raw observations. We could attempt to estimate means, variances and covariances between individuals using maximum-likelihood methods and then fit models in the usual way. This would involve using an unrealistically large number of parameters and prove extremely inefficient. The alternative approach is to formulate a model for these statistics and estimate its parameters directly from the raw data using a maximum likelihood method. Such procedures have been previously described (Elston and Stewart, 1971; Lange, Westlake and Spence, 1976). These authors show how to use the full likelihood of a set of pedigrees to obtain estimates of a set of parameters from the data. We will basically follow their approach, but employ more complex models in order to test hypotheses about cultural transmission and another numerical method.

We set up a vector of expected values for the means and an expected variance - covariance matrix for each family. The latter is equivalent to the model specified for the mean squares in Section 6. Then we choose a given set of parameter values which maximises the log

likelihood of observing the observed set of pedigrees. Lange, Westlake and Spence (1976) give a standard expression for the log likelihood of an observed pedigree, given a particular vector of means and an expected covariance matrix. Our main problem is to specify an expected covariance matrix i.e. a model. Lange, Westlake and Spence (1976) illustrate the method using simple models and Spence, Westlake and Lange (1976) have fitted the model to data on dermal ridge counts. They use expectations for simple situations with random mating and environmental differences within families. Additive genetical effects and dominance variation are included in the model. More complex models specifying genotype-environment interaction or genotype-environment covariation are not given. Since we are interested in the latter, we will describe more complex models in the following pages and show how the methods of Lange, Westlake and Spence (1976) may be used to fit them to our data.

2 SOLUTION OF THE NUMERICAL PROBLEM

2.1 Specification of the Log Likelihood

We shall follow Lange, Westlake and Spence (1976) in order to specify the likelihood function. They consider a pedigree of n people. In our case, n is the maximum possible pedigree size. Inspection of the data shows that this pedigree has $n = 6$ and is composed of the following members: mother, father, two adopted individuals and two natural siblings.

Let $\underset{\sim}{\mu} = (\mu_1 \ldots\ldots \mu_n)^t$ be a column vector of the expected values of the individuals of the pedigree, where $\mu_i = \mu_m$ or μ_f depending on whether the i^{th} individual is male or female. Males and females thus take separate means, μ_m and μ_f and these are the first two parameters of our model. Other models might be specified for the expected values.

Let $\underset{\sim}{x} = (x_1 \ldots\ldots x_n)^t$ be the observed trait values.

Finally, let $\underset{\sim}{\Omega}$ be a matrix of expectations for the variances and covariances between members of the pedigree, the expectations being expressed in terms of K parameters. Thus the total number of parameters in our model is (2 + K). The precise expectations for the terms of will be specified later.

Where we do not have a complete pedigree of n individuals (i.e. in most cases), missing members are ignored and $\underset{\sim}{\mu}$ and $\underset{\sim}{\Omega}$ reduced in size.

The natural logarithm of the likelihood of the pedigree is then given by:

$$L = -\tfrac{1}{2} \ln|\Omega| - \tfrac{1}{2}(x-\mu)^t \Omega^{-1} (x-\mu),$$

(Rao, 1973) (where $|\Omega|$ is the determinant of Ω).

The problem now is to find an appropriate method for choosing a set of values of the (2 + K) parameters which maximises this likelihood summed over families. Some version of Newton's method is most commonly used. This method requires that the first and second derivatives of the likelihood function are specified algebraically, or at least extracted numerically. Lange, Westlake and Spence (1976) produce an algebraic solution, which is complex even for their simple linear model. Since we shall specify far more sophisticated models during our investigation, we would prefer a method for maximising the likelihood which does not rely on the derivatives, or else uses numerical differentiation.

A second problem arises because the likelihood of observing a particular pedigree involves the inverse of the covariance matrix. Therefore, any set of parameter values which would yield a non-positive definite (NPD) covariance matrix will lead to failure of the method.

2.2 Maximising the Log Likelihood

Both these problems may be overcome by using a computer library routine, called E04HAF, supplied by the Nottingham Algorithms Group (NAG) (1977). This routine minimises a function of many independent variables, subject to general inequality and/or equality constraint functions, and bounds on the variables. Therefore, we use the program to minimise minus the log likelihood (i.e. to maximise the log likelihood). E04HAF uses iterative methods and requires initial estimates of the

position of the minimum as trial values. It can overcome the problem of a non-positive definite covariance matrix, if good trial values for a reasonable model are supplied. Otherwise the problem may be overcome by constraining the search for a minimum not to enter an area where $\underset{\sim}{\Omega}$ is NPD. One method of doing this which is often effective in practise is to use a penalty function technique which devises functions of the parameters which tend to zero when certain constraints are satisfied. The NAG routine uses a penalty function technique due to Lootsma (1972). A routine is supplied by the user to calculate the values of a number of problem functions for any set of values of the variables. To overcome our particular problem and ensure that the n x n covariance matrix is positive definite, we supply n functions, specifying that all the eigenvalues of the matrix are positive. However, this approach leads to a marked increase in computer time and is, therefore, undesirable. The necessity to specify such constraints may be overcome in most cases by the judicious choice of trial values or rectangular constraints on the parameter values.

The NAG routine provides a choice of three methods for the minimisation. One method, based on the Powell 64 method, relies on the function values only. It has no information on where to search for the solution in the space available and even with good trial values tries widely divergent values for the parameters before coming close to the solution. This means that, without the use of constraints, the covariance matrix often becomes NPD. The use of this method with constraints is extremely time-sonsuming and it's use is impracticable for our application. However, we did try it several times in this study for simple cases.

The second method uses values and first derivatives in a quasi-Newton method and the third also uses second derivatives in a Newton-type method. It is found, in practise, that the numerical differentiation used by these latter methods gives reasonably satisfactory results, (although it can be time consuming), as long as the iterative process is started with good trial values and a reasonable model is used.

The use of the NAG routine requires that a number of subroutines are supplied by the user. The program incorporating the E04HAF minimisation routine used in the following analyses was written by Dr. L.J. Eaves, Department of Genetics, University of Birmingham and was adapted for the particular problems to be described in the following pages. Further details of the program and methods will be described later when appropriate. All the computations were carried out on a CDC 7600 computer at the University of Manchester Regional Computer Centre via the SWAN link between Manchester and Birmingham's ICL 1904S and 1906A computers.

3 A SIMPLE APPLICATION OF THE METHOD

3.1 Comparison of Two Methods

The methods we are about to use are radically different from those employed in conventional model-fitting to twins, as described for the Georgia Twin Study and the Twin Study of Social Attitudes. Therefore, a comparison of the two approaches is highly desirable.

In theory, any discrepancies between the two methods should be due to inadequacies of the conventional approach through model-fitting by weighted least squares. Given normality of the original observations, the new approach should provide the maximum-likelihood solution of our problem, whereas weighted least squares is only equivalent to a maximum-likelihood method given certain assumptions. The most important of these from our point of view is normality of the distribution of the mean squares. Since we have shown that our data are approximately normally distributed and that the mean squares are based on quite large numbers, we do not expect greatly different answers from the two approaches.

The practical problem is somewhat different. The maximum-likelihood method is extremely complex compared with the older method. The NAG routine requires the user to specify the required accuracy of the solution. This is not easy since if the criterion is too stringent, convergence may never be reached. However, relaxing the criterion too far may produce biased parameter estimates. If the estimates from the weighted least squares and maximum-likelihood methods are closely similar, then we can suppose that our convergence criterion is adequate.

In order to compare the two methods, we shall attempt to fit the E_1D_R and E_1BD_R models to the twin data using the new method, before attempting to fit models to the twin and adoption data jointly.

3.2 Specification of the Data Structure and of Models

The concepts needed to fit our final complete genetical model will be introduced a few at a time, as they are needed to fit the simpler models that we shall consider first. The first two ideas we need to discuss are the specification of a matrix of relationships and of model matrices.

Before we can specify a model for the means, variances and covariances between relatives, we must uniquely identify each individual type and his relationship with all other members of his family. The following system has been devised for this purpose so that a complete pedigree can be drawn from a matrix of codes (Eaves, personal communication).

The diagonal terms of the matrix represent the individual members of the pedigree in the same order as their scores have been coded and the value of the diagonal term provides information about the individual. The off-diagonal terms contain codes specifying the relationships between all possible pairs of individuals. The system of coding to be used throughout is given in Table B35. An example of a pedigree illustrating the use of these codes is given in Table B36. Family 645 in our study demonstrates the use of all the codes, except that for an adopted son. From this matrix we can draw out the family pedigree without any further information. Eaves (personal communication) has extended this coding system so that pedigrees going back several generations can be specified.

TABLE B 35 : CODES USED IN SPECIFICATION OF PEDIGREES

	CODE	MEANING
Meaning of Diagonal terms	1	Male
	2	Female
	3	Adopted Male
	4	Adopted Female
Relationship specified by off-diagonal terms	20	Spouse of
	21	Natural child of
	22	Natural parent of
	23	Co-sibling of
	80	Adopted child of
	81	Adopting parent of
	82	Adopted sibling of natural child or viceversa
	83	Adopted sibling of an adopted child

TABLE B 36 : EXAMPLE OF A CODED PEDIGREE (FAMILY 645)

	ADOPTED DAUGHTER	MOTHER	FATHER	NATURAL SON	NATURAL DAUGHTER
ADOPTED DAUGHTER	4	80	80	82	82
MOTHER	81	2	20	22	22
FATHER	81	20	1	22	22
NATURAL SON	82	21	21	1	23
NATURAL DAUGHTER	82	21	21	23	2

From Table B36, we can see that the upper and lower triangles of the matrix yield the same information about the family and we could draw the pedigree from either. Thus, when we produce a data file to be used by the computer program, only the upper triangle of the relationship matrix is specified. The data file will contain the following information for each family in turn:

1. The number of individuals in the pedigree
2. The upper triangle of the relationship matrix
3. Scores for each individual (in the same order)

In order to specify a particular model, we now only have to give the expectations appropriate for each code. The program then calculates the expected covariance matrix for each family in turn using the information provided.

In order to generate the expectations for each twin family based on the E_1D_R model, as given in part A of Table B37, we simply attach values to the codes as shown in part B of this Table. In later Sections more complex models will be specified using all the data, but the principle for generating the expected covariance matrix for each family in turn remains the same.

3.3 Obtaining Maximum-Likelihood Solutions of the Parameters

Having specified the data and model, we now attempted to find the set of parameter values which would maximise the log likelihoods, summed over pedigrees, using the program incorporating the NAG library routine, E04HAF, for minimising a function. We have to provide a trial

TABLE 37 : EXPECTATIONS FOR THE E_1 D_R MODEL IN THE TWINS

PART A : Expectations for the Three Types of Twin Family.

	VECTOR OF MEANS	VARIANCE-COVARIANCE MATRIX
Monozygotic Twins	$\begin{pmatrix} \mu_m & \mu_f \\ \mu_m & \mu_f \end{pmatrix}$ or	$\begin{pmatrix} \frac{1}{2}D_R + E_1 & \frac{1}{2}D_R \\ \frac{1}{2}D_R & \frac{1}{2}D_R + E_1 \end{pmatrix}$
SAME-SEX DIZYGOTIC TWINS	$\begin{pmatrix} \mu_m & \mu_f \\ \mu_m & \mu_f \end{pmatrix}$ or	$\begin{pmatrix} \frac{1}{2}D_R + E_1 & \frac{1}{4}D_R \\ \frac{1}{4}D_R & \frac{1}{2}D_R + E_1 \end{pmatrix}$
OPPOSITE-SEX DIZYGOTIC TWINS	$\begin{pmatrix} \mu_m & \mu_f \\ \mu_f & \mu_m \end{pmatrix}$ or	$\begin{pmatrix} \frac{1}{2}D_R + E_1 & \frac{1}{4}D_R \\ \frac{1}{4}D_R & \frac{1}{2}D_R + E_1 \end{pmatrix}$

PART B : Specification for the codes

CODE	VARIANCE	MEAN
1	$\frac{1}{2}D_R + E_1$	μ_m
2	$\frac{1}{2}D_R + E_1$	μ_m
10	$\frac{1}{2}D_R$	-
11	$\frac{1}{4}D_R$	-
Other	-	-

estimate of minus the log likelihood which must be less than the final minimum achieved. Since this was the first run of the program, we had little idea of where the minimum would be. An estimate of -300.0 was given. This turned out to be far less than the actual minimum which was +311.6, corresponding to a log likelihood of -311.6. Convergence takes longer the further the estimate is below the true minimum. However, the program fails if the estimate is greater than the actual minimum. On this first run, a low estimate was used to prevent failure of the program, but on all later runs, more realistic estimates of the minimum were used in order to save computer time.

Trial values of the parameters must also be provided. These were taken from previous analyses of the data and were:

$$\mu_m = 0.005$$
$$\mu_f = -0.002$$
$$E_1 = 0.250$$
$$D_R = 1.000$$

Since, we expect these trial values to be close to the final parameter estimates, the problem of the search for a minimum entering a region where the expected variance-covariance becomes non-positive definite should be lessened. Therefore, we tried to obtain a solution using the function values only (i.e. the Powell 64 method). The program failed with the expected covariance matrix becoming NPD. In order to overcome this problem, we reran the program, constraining E_1 and D_R to be positive. This time a minimum was quickly reached. The parameter estimates and the corresponding likelihood are given in Part A of Table B38.

TABLE B 38 : COMPARISON OF MODELS FOR RADICALISM OBTAINED BY TWO DIFFERENT METHODS

PART A : THE E_1 D_R MODEL

Estimate	Maximum Likelihood Solution		Weighted least square solution
E_1	0.2536		0.2536
D_R	1.0072		1.0068
m	0.0266		-
f	0.0158		-
L	-315.6243	χ^2_8	14.276

PART B : THE E_1 B D_R MODEL

Estimate	Maximum Likelihood Solution		Weighted least squares solution
E_1	0.2689		0.2687
B	0.2093		0.2091
D_R	0.5753		0.5765
μ_m	0.0265		-
μ_f	0.0160		-
L	-311.6261	χ^2_7	6.198

The $E_1 B D_R$ model, which has been described previously, was similarly specified and a minimum was obtained using the function values only, by constraining E_1, B and D_R to be positive. The following trial values were used:

μ_m = 0.005

μ_f = -0.002

E_1 = 0.270

B = 0.210

D_R = 0.580

The parameter estimates and the corresponding maximum likelihood are given in Part B of Table B38. This Table also contains the weighted least squares solutions for both models. The estimates obtained by the two methods are closely similar, suggesting that reservations expressed earlier were ill-founded and that the two methods may be regarded as equivalent.

This shows that any problems of accuracy with respect to the maximum likelihood method used are minimal, and indicates its use for analysing the data of the Adoption Study. The comparison between the two methods also serves to provide some evidence concerning several criticisms that have been levelled at the use of weighted least squares for model fitting. One problem is that weighted least squares only provides maximum likelihood solutions when the mean squares are normally distributed. With data such as ours which are approximately normally distributed, no obvious differences are introduced by the use of weighted least squares. Another criticism has been that in using data summaries such as the

analysis of variance, information has been lost. Again, we find no evidence for this from the comparison of the two methods. From a practical point of view, both methods are essentially similar apart from the time taken to reach a solution and it turns out that the maximum likelihood method takes about a hundred times longer than the least squares method to reach covergence for these simple cases, where good trial values are available. With more complex models, or cases where no good trial values are available, time will be an even greater problem.

Model fitting by weighted least squares is clearly the most appropriate method for regular pedigrees where independent summary statistics such as mean squares can be obtained. When we have unbalanced pedigrees such as those of the Adoption study, weighted least squares cannot be used since we cannot easily obtain independent estimates of summary statistics and the equivalent, but more time-consuming method must be used. Before we consider twin and adoption data jointly, it is necessary to introduce another important concept which will be used throughout.

3.4 Hypothesis Testing

Weighted Least Squares has another advantage over the Maximum Likelihood method since it provides a well-defined and explicit test of the model. The residual variation after the model has been fitted follows the chisquare distribution. Since the properties of this distributinn are well-known, we can test whether the residual variation differs significantly from that expected by chance. We are unable, at present, to make such a test when the maximum likelihood method is used.

However, it is possible to compare alternative hypotheses about the observed variation, using the likelihood ratio criterion suggested by Elston and Stewart (1971). If H_o is one hypothesis about the data with a maximum log likelihood of L_o and H_1 is an alternative, more complex hypothesis with a corresponding likelihood of L_1, then the statistic, $LR = 2(L_1 - L_o)$, provides a test of the null hypothesis (H_o), since in large samples it follows an approximate chisquare distribution. The number of degrees of freedom is the difference in number of independent parameters between H_o and H_1 (Rao, 1973; Weiss, 1975). We can illustrate the use of the LR criterion by considering the two models we have already fitted. In Section 6, we saw that the difference between the chisquares for the E_1D_R and E_1BD_R models is itself a chisquare for one degree of freedom. This chisquare ($\chi^2_8 - \chi^2_7 = \chi^2_1 = 8.078$) was significant, indicating a significant reduction in the residual variation when the additional parameter is included in the model. The E_1BD_R model provides a significant improvement in fit over the E_1D_R model.

Twice the difference between the likelihoods of these two models yields a chisquare for one degree of freedom of 7.9964, which is not substantially different from the chisquare obtained in the weighted least squares analysis, again indicating that the additional parameter is accounting for a significant proportion of the observed variation. The difference between the χ^2_1's obtained using the two methods is less than 0.1, suggesting that the two methods are equivalent and that no serious loss of information occurs in the conventional weighted least squares analysis.

The likelihood ratio criterion can be used to compare alternative hypotheses about the data. But we would also like a test of the "goodness of fit" of particular models. The value of the log likelihood depends upon a number of factors other than the residual variation and thus cannot be used directly as a criterion for the adequacy of a model (such factors include the number of pedigrees entering into the analysis and the scale of measurement).

A likelihood ratio between a particular model and a model accounting for all the variation, would yield a chisquare, testing the significance of the residual variation after fitting the former model. The latter model might be specified by fitting a separate parameter for each mean, variance, and covariance. This is similar to the approach adopted in weighted least squares analysis. However, in practice, many problems are encountered which have yet to be overcome, and we have been unable to obtain the maximum likelihood solution for all these parameters.

There are two main reaons for this. The first is that the expected covariance matrix easily becomes non-positive definite. Therefore, we have to specify n constraints, where n is the number of unique relationships. Time to covergence is greatly increased using these constraints. Secondly the number of parameters involved is large even when the data structure is simple as in the case of the twin study. As the number of parameters is increased the time taken increases disproportionately. The upper limit on the number of parameters which can reasonably be fitted in this way, using constraints, appears to be about ten.

When we look at the complex structure of the adoption study, the problem is greatly multiplied. However, we made an unsuccessful attempt

to find a solution for the following ten parameter model: μ_m, μ_f one total variance, one correlation for both adopted-adopted and adopted-natural pairs of siblings and one correlation for each other type of relationship. This model would not account for all the variation if there are mean differences other than sex differences or differences in total variance between groups. Its maximum log likelihood might, therefore, be less than that of a model accounting for all the observed variation. Thus, the test of significance, given by $2(L_1 - L_0)$, would in some cases lead us to wrongly accept an inadequate model. However even such an imperfect test would be better than no test, especially since we have no evidence for significant differences means or total variances in our data.

An attempt to find the set of parameter values for this model (maximising the log likelihood summed over pedigrees) was made using all the methods available in E04HAF. The minimum number of constraints necessary were always used. Estimates of the parameters obtained from previous analyses were used as trial values. Yet despite all this and restarting the program several times with new trial values, we did not manage to achieve convergence.

Two other approaches for testing the model might prove more fruitful. We could attempt to specify our oversimplified model, equivalent to the 'E_1' model applied to twin data. For example, we might fit one mean, one total variance and no covariances. This would give some idea of the likely minimum value that the likelihood can take. Then we could at least say which of our models are significantly better than this.

The raw observations to which the model were fitted cannot be predicted from the parameter estimates (as we do in weighted least squares

in order to obtain a chisquare test). Spence, Westlake and Lange (1976) are trying to devise a method similar to the test used in weighted least squares. They take large pedigrees, and use the score of one individual in the pedigree as a baseline, in order to predict the scores of the remaining individuals and obtain a series of observed and expected scores for comparison. Even if a satisfactory test could be made in this way, it would not be very useful for our study, since we have very few pedigrees large enough.

There is an alternative to predicting the observations. We can calculate the covariances between the individuals in our study. We should also be able to predict these covariances from the parameters of our model. Thus it should be possible to devise a suitable test of significance.

However, in the following analyses, we have no test of "goodness of fit". Although in this thesis we are unable to test the adequacy of the models fitted by pedigree analysis, this problem will be returned to at a later date, since the method cannot be regarded as satisfactory until such a test has been found.

Another inadequancy of the pedigree analyses to be described in the following pages is the lack of standard errors for the estimates. This is not a theoretical problem, such as the specification of a test of "goodness of fit" which was described above. In order to obtain standard errors, we need to determine the values of the parameters which will produce a change of 0.5 in the likelihood. We could do this by numerical differentiation and obtain second derivatives. This assumes that the shape of the likelihood surface around the minimum is quadratic. Alternatively, we could plot the likelihood surface around the minimum.

In either case, the minimisation would have to be repeated, which would be expensive in time and money. The standard errors have not been found for the estimates of the parameters obtained in the analyses to be reported here. Standard errors for the parameters of a few of the more important models will be calculated at a later date.

4 MODELS FOR THE ADOPTION STUDY AND ALL DATA JOINTLY

Having shown that the method is satisfactory, we now wish to fit models to the adoption data and all the data jointly and compare these cases with the results from the twin study.

Expectations for all possible code values of the expected variance-covariance matrix are given in Table B39 for the simple E_1D_R model. Solutions for the parameters of the mean vector and covariance matrix were found for the adoption study (311 families) and then for all the data jointly (898 families), using function values only. The covariance matrix was constrained to be positive definite. The trial values used for the twins were again given. The results of fitting the model to these two data sets are given in Table B40, where the results from the twin study are also tabulated for comparison.

Firstly, we can make some qualitative comparisons of the parameter estimates. These are fairly similar in twin and adoption studies, but E_1 and D_R are both larger in the adoption study, reflecting the greater variance of these data (see Section 5). However, when we look at the genetical variation as a proportion of the total, we see that the additive genetical variation is very similar in the two data sets.

Parameter estimates for all the data jointly lie between those of the two studies, as expected, but are closer to those of the twin study, since there are more twin families (587, compared with 311). Of the 311 families of the adoption study, a large proportion are single individuals who contribute information only about the means and total variances, but not about the covariances.

TABLE B 39 : EXPECTATIONS FOR ALL CODES ON THE E_1 D_R MODEL

CODE	E_1	D_R
1,2,3,4	1	$\frac{1}{2}$
10	-	$\frac{1}{2}$
11,21,22,23	-	$\frac{1}{4}$
20,80,81,82,83	-	-

TABLE B 40 : MODEL FITTING IN ALL DATA JOINTLY, TWIN STUDY AND ADOPTION STUDY

	ALL DATA JOINTLY	ADOPTION STUDY	TWIN STUDY
μ_m	-0.02287	-0.12906	0.02656
μ_f	-0.00485	-0.03778	0.01570
E_1	0.25741	0.32999	0.25361
D_R	1.09520	1.14270	1.00720
$\frac{1}{2}D_R$	0.54760	0.57135	0.50360
h_N^2	0.68024	0.63389	0.66507
L	-517.9692	-198.4279	-315.6243

Now we will quantify our comparison of the two data sets by asking whether there is significant heterogeneity between them in respect to the parameters estimated. We cannot compare their log likelihoods directly since these are influenced by such factors as the number of pedigrees involved. We follow Lange, Westlake and Spence (1976) and let the parameter vector of the adoption study be γ_1 and of the twin study be γ_2. The null hypothesis, H_o, is that $\gamma_1 = \gamma_2$ and may be tested against the alternative hypothesis, H_1: $\gamma_1 \neq \gamma_2$, using the likelihood ratio statistic: $LR = 2(L_1 - L_o)$, which is approximately distributed as a chisquare with 4 degrees of freedom. L_1 is the sum of the likelihoods for the two studies considered separately and L_o is the likelihood for the two studies combined. It turns out that $LR = \chi^2_4 = 7.834$ ($5\% < p < 10\%$). Thus, there is little evidence for significant heterogeneity between twin and adoption studies, with respect to the four parameters estimated. If this model adequately explains the observed variation, then we have no evidence to suggest that the two data sets are not drawn from the same population, with common genetical and environmental causes of variation. The best estimates of the parameters are, therefore, those from the analysis of the combined data. However, although formally the E_1D_R model is adequate to explain the observations in the twins, the fit is poor. The addition of a between families component of variation produces a significant reduction in residual chisquare and the E_1BD_R model fits the observations closely. Since $\hat{B}$ is significant in the twins, and the adoption data are more variable than the twin data, in order to be confident that the data are homogeneous, we need to compare the two studies with respect to the parameters of the E_1BD_R model.

5 BETWEEN FAMILIES VARIATION IN THE TWO STUDIES

Adoption data provide a unique opportunity to study components of between families variation. Ultimately, we wish to consider the roles of assortative mating and cultural differences in producing this variation. But in the first instance we will fit a model analagous to the E_1BD_R model in the twins, in order to determine whether the between families component is homogeneous over twin and adoption data.

Specification of the between families component presents problems because we now have intergenerational as well as collateral relationships. In order to formulate a model, we assume no assortative mating and set the expectation for the covariance between spouses to zero. We are thus assuming that variation estimated as $\hat{B}$ in the twin data is produced only by environmental differences between families (E_2). This assumption will be tested later, when we compare the maximum log liklihoods of models with no assortative mating or E_2, with either assortative mating or E_2 only and with both effects.

A number of models for the environmental covariance between individuals living together are possible, but we will first consider the model most similar to that fitted to the twins (i.e. E_1D_RB). This model assumes that the environmental factors producing variation in the children are independent of parental phenotype, that the environmental covariance arises anew in each generation and that the environment acts to produce variation in the children. Thus, there is an environmental covariance between collaterals raised in the same home, E_2, which is added to the expectations for the total variance and all collateral covariances to produce the E_1D_RB model given in Table B41. This model for the

TABLE B 41 : EXPECTATIONS FOR A MODEL WITH A COMMON ENVIRONMENT FOR COLLATERALS

CODE	E_1	D_R	E_2
1,2,3,4	1	$\frac{1}{2}$	1
10	-	$\frac{1}{2}$	1
11,23	-	$\frac{1}{4}$	1
20,80,81	-	-	-
21,22	-	$\frac{1}{4}$	-
82,83	-	-	1

environmental variation between families, which is similar to that used in the twins, makes no additional assumptions about the structure of the environment. Therefore, it can be used to determine whether differences between additive genetical or environmental components in the twin and adoption data are significant.

Estimates of $\hat{E}_1$, $\hat{D}_R$ and $\hat{B}$ from the twins were used as trial values. The program was run using the function values and constraining the covariance matrix to be positive definite. The solution improved very slowly for both the adoption data alone and the combined data. When the time limit was reached the program was rerun with the current solutions as trial values, using the method based on first derivatives. Convergence was quickly reached with the new method. The constraints were removed and the program was rerun using the final solutions as trial values, so that we could compare constrained and unconstrained solutions. The E_1D_RB model was also rerun in the twins without constraints. Convergence was quickly reached for the combined data and the twin data, but the covariance matrix quickly became non-positive definite for the adoption data. Results for constrained and unconstrained solutions are given, where possible, in Table B42. These are identical to the first six significant figures for the twin study. There is no unconstrained solution for the adoption study. Constrained and unconstrained solutions are somewhat different for the combined data. Unconstrained values of $\hat{E}_1$ and $\hat{E}_2$ are somewhat smaller and $\hat{D}_R$ is correspondingly larger. Since the total variance is the same for both solutions, the narrow heritability is larger for the unconstrained case. Its maximum log likelihood is also larger.

TABLE B 42 : COMPARISON OF CONSTRAINED AND UNCONSTRAINED SOLUTIONS OF THE $E_1D_RE_2$ (or E_1D_RB) MODEL IN THREE DATA SETS

ESTIMATE	ALL DATA JOINTLY		ADOPTION DATA		TWIN DATA	
	CONSTRAINED	UNCONSTRAINED	CONSTRAINED	UNCONSTRAINED	CONSTRAINED	UNCONSTRAINED
μ_m	-0.0203	-0.0214	-0.1301	-	0.0265	0.0265
μ_f	-0.0060	-0.0060	-0.0395	-	0.0160	0.0160
E_1	0.2931	0.2681	0.2288	-	0.2689	0.2689
D_R	0.5967	0.6779	1.2422	-	0.5753	0.5753
E_2(or B)	0.2225	0.2061	0.0526	-	0.2093	0.2093
L	-514.2221	-513.6002	-197.9073	-	-311.6261	-311.6261
V_T	0.8140	0.8132	0.9025	-	0.7659	0.7659
$\frac{1}{2} D_R/V_T$	0.3665	0.3669	0.6882	-	0.3756	0.3756
E_2/V_T	0.2733	0.2532	0.0583	-	0.2733	0.2733
E_1/V_T	0.3601	0.3297	0.2535	-	0.3511	0.3511

Constrained and unconstrained solutions in the twins were closely similar and we know from the weighted least squares analysis that the E_1D_RB model is adequate to explain variation in the twins. We have no test of the adequacy of the $E_1E_2D_R$ model in the adoption data and combined data. Problems such as slow convergence, differences between constrained and unconstrained solutions or failure to obtain an unconstrained solution might be warning us that the model is inappropriate and maximum likelihood solutions of its parameters are difficult to find. They might equally indicate that we have not chosen the most appropriate method for the particular problem. In either case, caution must be exercised when interpreting the results. In particular, unconstrained solutions will be sought whenever possible since the constrained solution may be less likely and give different parameter estimates.

Comparison of the solutions for the three data sets shows that the total variances of the adoption data, twin data and combined data differ but are similar to those estimated from the data in Section 5 (i.e. $V = 0.911$, 0.761 and 0.803 respectively). Thus, the model cannot be completely inappropriate. A brief look at the estimates from the twin and adoption studies reveals quite substantial differences. Before considering these in detail, we may ask whether there is significant heterogeneity between the two studies with respect to the parameter set as a whole. Using the likelihood ratio criterion described earlier, we find that LR = $\chi^2_5 = 9.3774$ ($5\% < p < 10\%$), based on the likelihoods from the constrained solutions. If we base the likelihood ratio on unconstrained solutions for twin data and combined data and the constrained solution for the adoption data, $\chi^2_5 = 8.1336$ ($10\% < p < 20\%$).

Neither chisquare provides evidence for significant heterogeneity between twin and adoption data. This is interesting in the light of criticism of the twin method which suggests that twins are special and may be subject to different environmental influences than single born children. There is no evidence for such differences between twins and other individuals in this study.

We may now ask what are the likely causes of differences in parameter estimates between twin adoption studies, bearing in mind that we only have the constrained solution for the adoption data, which might indicate that the model is inappropriate. Estimates of $\hat{E}_1$ are similar, although slightly larger in the twins. However, both $\hat{D}_R$ and $\hat{E}_2(B)$ differ markedly. $\hat{D}_R$ is much greater in the adoption study, accounting for 69% of the total variation, whereas $\hat{E}_2$ accounts for only 6%. In the twin data, $\hat{B}$, which provides a biased estimate of E_2, accounts for 27% of the variation, the contribution of $\hat{D}_R$ being reduced to 38% of the total variation.

Comparing estimates from twin and adoption data with those for the combined data, we see that estimates for the combined data closely resemble those of the twin study, suggesting that the twin study is providing most of the information about these parameters. Using the present model, information about environmental covariance comes only from pairs of collaterals in the adoption study, and there are relatively few of these, whereas parent-offspring and sibling pairs, of which there are many more, provide information about D_R. Thus, we may have a more powerful test for D_R than E_2 in the adoption study. This would be discouraging since we hope to partition the between families component using the adoption data. However, the present model is unrealistic

because we have specified no covariance between spouses (either genetical or environmental) and no environmental covariance between parents and the children they rear. Since we have data on a substantial number of spouses, natural parent-offspring and adopting-parent offspring pairs who would provide information about the family environment on a nore realistic model, a detailed study of the between families component of variation may not be precluded.

The discrepancy between the two data sets is most marked when we consider the between families component. In the twin data, we were estimating $\hat{B}$, which includes both common environmental effects and additive genetical variation produced by assortative mating. In the adoption data, we estimated only $\hat{E}_2$. Therefore, the difference between the two estimates, one of $\hat{B}$ and one of $\hat{E}_2$, may be due to assortative mating. This seems likely because we know there is a marital correlation, which we have disregarded in performing this analysis.

Another possible reason for differences between the two data sets is that we have used the wrong model for the environmental variation. This model may be adequate to explain the twin data, but not the adoption data where we have intergenerational relationships. We have allowed no environment covariance between parents and children reared by them.

Another model of the environment will now be considered which may be more realistic in light of the known correlations between individuals (see Section 5.3). We assume that environmental covariance arises between individuals, just because they are living together in the same household. Thus we add an E_2' term to the covariance of all individuals of one family. This model is shown in part A of Table B43. There is

TABLE B 43 : EXPECTATIONS AND SOLUTIONS FOR A MODEL WHERE WHOLE FAMILIES SHARE A COMMON ENVIRONMENT

PART A EXPECTATIONS

CODE	E_1	D_R	E_2^1
1,2,3,4	1	$\frac{1}{2}$	1
10	-	$\frac{1}{2}$	1
11,23	-	$\frac{1}{4}$	1
20,80,81	-	-	1
21,22	-	$\frac{1}{4}$	1
82,83	-	-	1

PART B SOLUTIONS

	ALL DATA JOINTLY	ADOPTION DATA	TWIN DATA
μ_m	-0.0312	-0.1592	0.0265
μ_f	-0.0042	-0.0436	0.0160
E_1	0.2679	0.3668	0.2689
D_R	0.7267	0.6831	0.5753
E_2^1 (or B)	0.1794	0.1911	0.2093 ($\hat{B}$)
L	-509.8290	-193.8480	-311.6261
V_T	0.8107	0.8995	0.7659
$\frac{1}{2}D_R/V_T$	0.4482	0.3797	0.3756
E_2^1/V_T	0.2213	0.2125	0.2733
E_1/V_T	0.3305	0.4078	0.3511

now environmental similarity between spouses, parents and children in one family and collaterals living together.

This model was fitted using the method based on first derivatives, with solutions from the previous model as trial values. Convergence was quickly achieved for the combined data and the adoption data, without the necessity to supply constraints. Estimates of the parameters and the maximum log likelihoods are given in part B of Table B43. The solutions for the E_1D_RB model in the twins are again given for comparison.

This model provides a substantially more likely solution than previous models, whether we consider the combined data or only the adoption data. However, we have no degrees of freedom for testing the significance of this improvement. In order to determine the source of the improvement, a similar model was fitted in which we specified environmental covariances between collaterals and between parents and children, but none between spouses (E_2''). A solution was quickly reached for this model, using first derivatives with no constraints and the same trial values as before. The solutions for the combined data and for the adoption data are given in Table B44. The estimates are fairly similar to those for the previous model given in Table B43. However, the likelihoods are substantially lower and are about the same as those based on the model with a single E_2 for collaterals only. Thus, the improvement of the $E_1D_RE_2'$ model over the $E_1D_RE_2$ model is undoubtedly due to the addition of a covariance between spouses.

Using the likelihood ratio criterion to provide a measure of the heterogeneity of twin and adoption data with respect to the parameters of the $E_1D_RE_2'$ model, we find that $\chi^2_5 = 8.7098$ $(10\% < p < 20\%)$.

TABLE B 44 : SOLUTIONS FOR A MODEL WITH ENVIRONMENTAL COVARIANCE BETWEEN COLLATERALS AND PARENT-OFFSPRING PAIRS

	ALL DATA JOINTLY	ADOPTION DATA
μ_m	-0.0291	-0.1458
μ_f	-0.0052	-0.0401
E_1	0.2664	0.4674
D_R	0.8327	0.6285
E_2^{11}	0.1255	0.1185
L	-514.3745	-197.1571
V_T	0.8083	0.9002
$\frac{1}{2}D_R/V_T$	0.5151	0.3492
E_2/V_T	0.1553	0.1316
E_1/V_T	0.3296	0.5192

Again the data are homogeneous with respect to this parameter set. However, the magnitude of the parameters has altered somewhat and the values for the adoption study are much closer to those of the twin study. $\hat{D}_R$ now accounts for only38% of the variation in the adoption data, rather than 69%. This is very similar to the narrow heritability in the twins. E_1 accounts for a similar proportion of the total variation in the two groups but the percentage is somewhat greater in the adoption data (41%) than the twin data (35%). This is because $\hat{E}_2'$ is still slightly smaller than $\hat{B}$ in absolute size and accounts for only 22% of the variation whereas $\hat{B}$ accounts for 27%. However, the differences between the parameters are now small compared with the previous $E_1D_RE_2$ model, suggesting we now have a more appropriate model for the environment. However, it is also possible that this model is the most likely so far, simply because we have allowed a covariance between spouses. Thus, the problem of discrimination between cultural or other environmental differences between families and assortative mating has still not been resolved. In order to make this resolution and to test for cultural transmission, we must consider another method of specifying models. Eaves (1976b) model of cultural transmission lets parental phenotype contribute to offspring's environment, leading to cultural differences between families. His paper gives all the expectations we need for our covariance matrix in conventional biometrical-genetical terms. However, his model does not specify the contribution of assortative mating. The extension of Eaves model to include assortative mating is likely to prove a long and difficult task and this has led us to explore another way of specifying the effects of cultural transmission and assortative mating

simultaneously, which may prove less algebraically demanding than the approach of Eaves and yield simpler expectations. The methods of path analysis which provide a visual representation of the relationships between the genotypes and environments of different individuals, are discussed in the next section.

SECTION 8: AN ALTERNATIVE APPROACH THROUGH THE USE OF PATH MODELS

1 INTRODUCTION

In previous sections, we have used the analysis of variance approach in order to determine the sources of individual differences in several traits, by fitting models to variance components. This methodology was initiated by Fisher (1918) and its use for solving problems in biometrical genetics extended by Mather and Jinks (see Mather and Jinks, 1971). An alternative approach to the problem was taken by Sewall Wright, who treated genetical problems purely as regression problems. The details of his approach are expounded in Wright's books, "The Method of Path Coefficients" and "Systems of Mating" and also in Li's "Population Genetics". Rao, Morton and Yee (1974, 1976) use path models for the analysis of family resemblance and their methods are appropriate for human data. Details of the methods can be found from any of these sources and only the aspects of path analysis necessary for the present study will be discussed as the need arises.

The path model approach suffers from a number of limitations which restrict its use especially in experimental organisms. When we are dealing with man, path models may be useful and the features of the approach which are usually found to be limiting may be put to good use.

The method normally requires two assumptions which we are not willing to accept without some test. Most workers use standardised paths for their models. This introduces all the problems associated with analysing corrolation coefficients, which were discussed in Part A.

Use of standardised paths assumes that the total variances of the groups being studied are homogeneous. If this assumption is not justified, biased and misleading estimates of the paths may be obtained. However, the bias can often be found, providing some information from another source is available (e.g. the correlation of parents and natural children not reared by them in the case of genotype-environmental covariance). Since we are interested in cultural differences and cultural transmission, which may lead to genotype-environment covariance, the use of standardised paths is clearly inappropriate, since we do not have such additional information. Eaves (1967a, b) has shown that two mechanisms producing genotype-environment covariance i.e. cultural transmission and sibling effects, both lead to expectations for the total variances which are not the same in all groups. We will derive expectations for the expected covariance matrix using a path model, and, therefore, we will not be using standardised paths.

Another problem of path analysis is that it is essentially a regression method and, therefore, we are unable to say much about the nature of the genes operating. Path models normally assume that all genetical variation is additive, since non-additive effects cannot be specified using path analysis. In contrast, the biometrical-genetical approach provides explicit tests for non-additive effects. Specification of such effects as dominance, epistasis and genotype-environment interaction is possible, and their effects can, therefore, be estimated, given an adequate experimental design.

The chief advantage of path analysis is that, since it is a regression technique, specification of effects at a phenotypic level

is easy. Thus biometrical-genetics is generally superior for analysing genetical variation, whereas the use of path models for analysing environmental variation may be advantageous. This is why we intend to use path models for analysing the between families component of variation in these data. The E_1BD_R model adequately explained variation in the twin data, and there was no evidence to suggest genetical non-additivity. Failure to detect non-additivity in the twin data is not surprising. Eaves (1972) showed that the power of the test for detecting dominance with twin data is very low with the sort of numbers used in this study and also that dominance tends to be counter balanced by E_2. However, it is possible for dominance to be non-significant in the twins but highly significant in family data where the test is more powerful, leading us to different conclusions in the two sorts of data. Eaves (personal communication) has shown in several instances that inconsistencies between twin and family data may be resolved by introduction of a non-additive component into the model even when there was no evidence for significant non-additivity in the twins alone. He found that estimates of dominance in twin and family data were similar in magnitude, but only significant in family data. However, in the present study no evidence was found for significant heterogeneity between twin and adoption studies and we have no reason to suspect that the genetical system is other than simple and additive. Even if there were some non-additive variation for social attitudes, we have no reason to suppose that we will be able to detect it in these data.

However, we do know that there is significant between families variation due to common environmental effects and/or assortative mating. We shall thus use path analysis to detect and estimate the principal

additive components of the system empirically. We found no evidence for genotype-environment interaction in the twin data. Therefore, the environment seems to be acting additively at the phenotypic level. Assortative mating is assumed to act through the phenotype. The association of like phenotypes leads to a covariance between the genetical similarity of spouses. Since this produces only additive genetical variation, we may specify the effects of assortative mating using path models.

We have some evidence for cultural differences between families and wish to explore the possible role of cultural transmission for social attitudes. Eaves' (1976b) model of cultural transmission gives expectations for variances and covariances expressed in biometrical-genetical terms. However, his model does not include the effects of assortative mating. We have evidence for assortative mating, since we know there are significant marital correlations for both Radicalism and Toughmindedness. Therefore, we wish to fit a model including cultural transmission and assortative mating simultaneously. The extension of Eaves' model to include assortative mating is likely to be a long and difficult task, since the formulation of the basic cultural transmission model was numerically and algebraically complex. In Eaves (1976b) model, the phenotype of the parent contributes to the environment of his offspring, producing variation between families. Since this model is based on phenotypes, the path approach may prove rewarding and allow the formulation of a model for cultural transmission and assortative mating simultaneously. Such a model is likely to prove simpler and algebraically less demanding than attempts reformulate Eaves' model of cultural transmission to include assortative mating.

2 TWO SIMPLE MODELS

Since the concepts and methods of path analysis have not been discussed before, there are many new ideas to be introduced. We shall, therefore, consider a series of simple and instructive examples, in order to describe the methods to be used in later Sections, when we will fit a full model including cultural transmission and assortative mating. These examples will also be useful in understanding the between families variation.

We shall first consider two extremely simple models to show the use of variances and correlations to specify models for all the elements of the variance-covariance matrices and how we use the maximum likelihood method to obtain solutions for the parameters. We do not need to introduce the concept of path models at this stage. These first two models are unrealistically simple and will, therefore, give some idea of the lowest values the maximum log likelihood can take.

In the first model, we assume the absence of genetical variation, but introduce an environmental covariance between individuals living in the same family. The population variance will be called "V" in this and all subsequent models. The degree of resemblance between individuals produced by the environmental covariance is given by a correlation, r, which is the same for all possible relationships in a family. The diagonal terms of the expected covariance matrix are all V, since the total variances for all individual types are expected to be the same. The covariances may be specified as a product of the variances and the correlation since

$$r = \frac{\text{Cov } x_1 x_2}{\sqrt{Vx_1 \; Vx_2}}$$

which gives the covariances as rV for all the off-diagonal terms.

Solutions for the four parameters of this basic model were found using the maximum likelihood approach described earlier. Convergence was quickly reached for both Radicalism and Toughmindedness, using the method based on first derivatives, without the necessity to use any constraints. Estimates of the parameters and the maximum log likelihoods are given in Table B45 for both traits. Estimates of V are the same as the total variances for the combined data, calculated in Section 5 (i.e. V = 0.803 for Radicalism and V = 0.082 for Toughmindedness in Section 5). For both traits the overall correlation between relatives lies between 0.4 and 0.5. We know this model is unrealistic because we saw in Section 5 that the monozygotic twin correlations (r_{MZ} = 0.62 for Radicalism and r_{MZ} = 0.60 for Toughmindedness) are higher than the DZ correlations (r_{DZ} = 0.49 for R and r_{DZ} = 0.46 for T), suggesting either greater genetical similarity between MZ twins or else that they are treated more alike. This model is far less likely (L = -533.7704) than the E_1D_R model (L = -517.9692), fitted to the data earlier, and we know that the E_1D_R model is barely adequate to explain the observed variation. This model must be approaching the worst model possible to explain the observations.

A simple extension to the model allows for the observation that MZ twins are more alike than DZ twins. A new parameter, t, is introduced which is the correlation between MZ twins introduced if MZ twins are treated more alike than other siblings. The only expectation which is affected is the covariance for monozygotic twins which is now given by:

$$\text{Covariance}_{MZ} = V(r + t)$$

TABLE B 45 : SIMPLE ENVIRONMENTAL MODEL FOR RADICALISM AND TOUGHMINDEDNESS FITTED TO ALL DATA JOINTLY

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0359	0.0077
μ_f	-0.0012	0.0054
V	0.8038	0.0822
r	0.4837	0.4375
L	-533.7704	1292.5851

We will call "t" the treatment effect, since it affects only the MZ twin correlation. However, since twins provide most of the information, t could also be expressing the greater genetical similarity between MZ twins compared with DZ twins, although we have not given sensible expectations based on genetical theory for the other relationships.

This model was fitted to the Radicalism and Toughmindedness data using the method based on first derivatives. Convergence was quickly reached without the use of constraints. The parameter estimates and maximum log likelihoods for the two traits are given in Table B46. The correlation for individuals reared in the same family is now reduced to 0.3711 for R and 0.3075 for T. The MZ correlations are boosted by the treatment effect to 0.6653 and 0.6162 respectively (i.e. r + t). The likelihoods for this model are significantly greater for both traits than those obtained when we estimated a single correlation between relatives.

The two parameter sets just estimated, are purely empirical data summaries, since they do not attempt to explain the variation using a theoretical framework. The second "model", however, which includes an additional correlation between MZ twins which might be predicted as a treatment difference, provides a more adequate summary of the data than the first model, indicating that the resemblance between individuals cannot be adequately represented by a single correlation. It is also somewhat more likely than the E_1D_R model, indicating the inadequacy of the latter, theoretical model to explain the observed variation.

The results found for Radicalism so far will now be summarised since this is likely to be the most interesting trait, when we fit

TABLE B 46 : ENVIRONMENTAL MODEL WITH TREATMENT EFFECTS

PARAMETERS	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0341	0.0081
μ_f	-0.0019	0.0063
V	0.8112	0.0823
r	0.3711	0.3075
E	0.2942	0.3087
L	-515.8289	1309.0122
r + t	0.6653	0.6162

models specifying genotype-environment covariance. The models, their parameters and corresponding maximum log likelihoods are given in Table B47.

The first model is clearly not an adequate representation of the data, since it is far less likely than any of the other models. If we wanted a four parameter summary of the data, the E_1D_R model provides a more likely explanation, suggesting that the degree of similarity between individuals depends on the closeness of their relationship i.e. there is a genetical component of variation. However, the E_1D_R model does not account for all the variation since the three 5 parameter models are all significant improvements over the 4 parameter models. The third model is significantly more likely than either of the first two models ($\chi^2_1 = 35.3950$ and $\chi^2_1 = 4.9308$, respectively). This suggests that we need a model including both common environmental effects and a greater similarity in MZ than DZ twins. Both model 4 and model 5 are of this type and are more likely than any other model considered so far, but we cannot test these models against the third model or one another since they all have the same number of parameters, although we see by inspection that Model 5 is the most likely. However, as predicted from the twin study, there is additive genetical and probably common environmental variation.

A comparison of models 5 and 6 suggests that covariance between spouses is also an important factor, and may well account for a substantial proportion of the between families variation.

We wish to investigate the nature of the common environmental variation further, in conjunction with assortative mating, and hopefully

TABLE B 47 : SUMMARY OF MODELS FITTED TO RADICALISM USING PEDIGREE ANALYSIS (BUT NOT PATH MODELS)

DESCRIPTION OF MODEL	NUMBER OF PARAMETERS	PARAMETERS	L
Simple model with one covariance between individuals living together.	4	$\mu_m \mu_f$ V r	-533.7704
Simple additive genetical model.	4	$\mu_m \mu_f$ E_1 D_R	-517.9692
Simple model with treatment effect.	5	$\mu_m \mu_f$ V r t	-515.8289
Simple additive genetical model with environmental covariance between collaterals.	5	$\mu_m \mu_f$ E_1 D_R E_2	-513.3635
Simple additive genetical model with environmental covariance between individuals living together.	5	$\mu_m \mu_f$ E_1 D_R E_2^1	-509.8290
Simple additive genetical model with environmental covariance between collaterals and parent-child pairs only.	5	$\mu_m \mu_f$ E_1 D_R E_2^{11}	-514.3745

find a model which is significantly more likely than any of those in Table B47. Models may be more likely which allow for environmental covariance between spouses and/or covariance between their genetical deviations. We shall attempt to specify such models using the approach of path analysis.

3 SIMPLE ENVIRONMENTAL PATH MODELS

We introduced, in the last section, the idea of specifying a model for an expected covariance matrix in terms of correlations and a variance, so that an off-diagonal covariance term is simply the product of the total variance and the correlation between the pair of individuals concerned. Such methods were fitted by maximising the log likelihoods, summed over pedigrees. We now want to use the path approach to specify models for the expected covariance matrix in a similar manner and fit these models by maximum likelihood. Thus, we shall not encounter the problems associated with standardised paths, and we will be able to allow the total variances of different groups of individuals to differ.

We shall first look at path models based purely on environmental effects, to determine whether any of these is more likely than the models already considered. This may tell us more about the structure of the environmental covariance and will give usabaseline against which to compare the likelihood of the genetical path models to be introduced later. We wish to consider environmental covariance between relatives and assortative mating simultaneously. We can do this, assuming no genetical variation, if the assortative mating is based on the phenotype. We know that there is a marital correlation. However, we will take this opportunity to determine whether inclusion of a marital correlation provides a significant improvement over simple environmental models, with no marital correlation.

In order to formulate the models, we will make the following assumptions:

1. The total variances of all groups are equal
2. There is no genetical similarity between relatives
3. All children are treated similarly by the parents who rear them regardless of whether they are natural or adopted, twins or single births.

Two basic models, with and without assortative mating, based on the assumptions stated above are given in Figure B5. The first model implies that similarity between parents and children reared by them is produced by the equal environmental impact of the phenotypes of both parents, by the path "p". This model is extremely simple and might be interpreted in other ways.

The second model includes the direct environmental influence of parental phenotype on the offspring through the path "p", but we now specify an environment, E, which is common to all collaterals, but independent of the parents. The path from E to offspring phenotype is called "e".

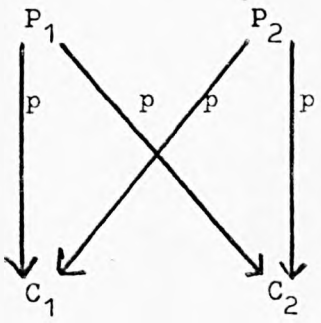
Expectations for parent-offspring, sibling and marital correlations can now be determined using the conventional rules governing paths (Li, 1955). For example, in order to determine the correlation between the two children in model 2B, we can trace back from C_1 to E along path e and then forward along path e from E to C_2 to give us a correlation of e^2. Tracing paths from C_1 to C_2 via P_1 gives a correlation of ρ^2. Similarly another term in ρ^2 is obtained from the paths from C_1 to C_2 via P_2. Correlations of $\rho^2 \mu$ are produced by tracing paths: 1. From C_1 to P_1 to P_2 to C_2. 2. From C_1 to P_2 to P_1 to C_2. Adding all the terms, the total correlation becomes:

$$r_{siblings} = 2\rho^2 + 2\rho^2\mu + e^2$$
$$= 2\rho^2(1 + \mu) + e^2$$

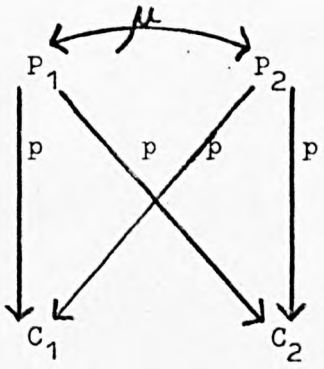
The total variance, as used earlier, is V. Therefore, the covariance

FIGURE B5: ENVIRONMENTAL PATH MODELS

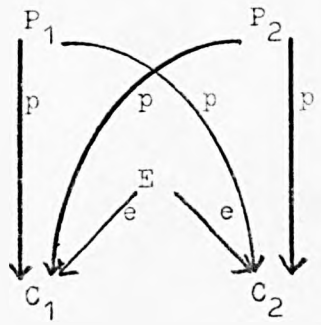
1A.



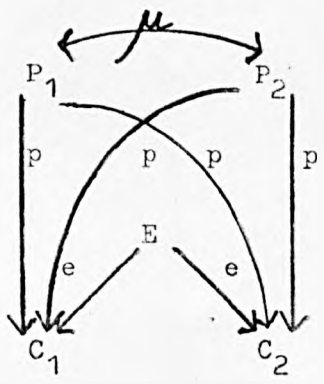
1B.



2A.



2B.



KEY:

P_1 First parent

P_2 Second parent

C Child

E Environment (see text)

p, e } Paths (see text)

μ Marital correlation

between siblings is:

$$Cov_{siblings} = V(2^2 (1 + \mu) + e^2)$$

The covariances for all four models were calculated and these expectations are given in Table B48. Solutions for the parameters of the four models, which maximised the log likelihood summed over pedigrees were found using the methods described earlier. Estimates of 500.0 and -1500.0, for Radicalism and Toughmindedness respectively, were given for minus the log likelihood. Earlier analyses led us to believe that the final minimum minus the log likelihood would not be below these values and produce failure of the program. The actual values of male and female means, total variances and marital correlations calculated in Section 5 were given as the trial values. Values of 0.5 were used as trial values for the paths, p and e, since we had no idea about their likely magnitude. The covariance matrix was constrained to be positive definite because the trial values for the paths were unlikely to be very realistic. Solutions for models 1A and 1B were obtained using the function values only. The method based on the first derivatives was used to obtain the maximum likelihood solution for models 2A and 2B. Estimates of the parameters and their corresponding maximum log likelihoods are given in Table B49 for Radicalism and Toughmindedness.

Model 1 is clearly inappropriate for these data for both Radicalism and Toughmindedness, whether the marital correlation is included or not, since the likelihoods are substantially lower than those of any of the other models we have considered so far, even the model where only one correlation is specified to account for resemblance between individuals living together in the same family. It is, therefore, not even a useful summary of the data.

TABLE B 48 : EXPECTATIONS FOR SIMPLE ENVIRONMENTAL PATH MODELS

STATISTIC	CODES	MODEL 1A
Total variance	1,2,3,4	V
Parent-offspring covariance	21,22,80,81	pV
Sibling covariance	10,11,23,82,83	$2p^2V$
Covariance of spouses	20	0

EXPECTATIONS

MODEL 1B	MODEL 2A	MODEL 2B
V	V	V
$pV(1+\mu)$	pV	$pV(1+\mu)$
$2p^2V(1+\mu)$	$V(2p^2+e^2)$	$V(2p^2(1+\mu)+e^2)$
$V\mu$	0	$V\mu$

TABLE B 49 : PARAMETER ESTIMATES FOR SIMPLE ENVIRONMENTAL PATH MODELS

PARAMETER	ESTIMATES							
	RADICALISM				TOUGHMINDEDNESS			
	1A	1B	2A	2B	1A	1B	2A	2B
μ_m	-0.0409	-0.0436	-0.0218	-0.0224	0.0047	0.0073	0.0090	0.0108
μ_f	-0.0048	-0.0002	-0.0066	-0.0056	0.0027	0.0047	0.0056	0.0067
V	0.8100	0.8126	0.8100	0.8167	0.0824	0.0829	0.0827	0.0828
p	0.4752	0.4175	0.1417	0.1335	0.4202	0.3564	0.0052	0.0155
e			0.7200	0.7132			0.7302	0.7299
μ		0.3278		0.3158		0.4662		0.4224
L	-553.7629	-547.4797	-523.4731	-519.4698	1263.5794	1271.9892	1310.6675	1314.7409

Model 2 is more likely than model 1 for Radicalism, but still provides a worse explanation of the data than either the simple environmental model with treatment effects or either of the simple genetical models with common environmental variance. Even Model 2B which includes a marital correlation and is, therefore, a six parameter model, is less likely than any of the five parameter models. This shows that our model is totally inadequate. The problem obviously arises because we have not allowed for closer relatives to be more alike, which we do in all the five parameter models. This reinforces the findings from previous analyses that a genetical component is needed to explain variation in Radicalism.

In the case of Toughmindedness, Model 2A gives a maximum log likelihood of 1310.6675, which is slightly larger than that for the environmental model with treatment effects (ML = 1309.0122, see Table B46). This model is unrealistic because we know there is a correlation between spouses. Model 2B, which includes a marital correlation provides a significant improvement over the environmental model with treatment effects ($\chi^2_1 = 11.4574$, $p < 0.001$). Therefore, for Toughmindedness, it seems that we may not need to allow closer relatives to resemble one another more closely, which agrees with our finding from the twin study that there is little evidence for a genetical component. Model 2B is a significant improvement over model 2A. Therefore, the most likely model from the evidence we have so far is that resemblance between the individuals in our study can be explained in terms of assortative mating between spouses, based on their phenotypes, an environment which is mediated through the phenotype of the parents and also a common environmental component independent of parental phenotype.

We may now ask what the actual values of the parameters tell us. We will consider only the parameters of model 2B, since this is the most likely model for both traits. The values of the total variances are close to those calculated from the data in Section 5, as we expect. However, the marital correlations, $\mu = 0.3148$ and $\mu = 0.4224$ for Radicalism and Toughmindedness respectively are not quite what we expect since we found $\mu = 0.4045$ and $\mu = 0.3824$ for these traits in Section 5. The discrepancy is greater for Radicalism than Toughmindedness and indicates that Model 2B is not accounting for all the variation, particularly in the case of Radicalism.

The relative magnitudes of p and e are quite informative. For both traits, e is substantially larger than p, suggesting that an environmental covariance not dependent upon the parents is of some importance. The value of p is quite large for Radicalism ($p = 0.134$), but is negligible for Toughmindedness ($p = 0.015$). This suggests that the common environment producing similarity in Toughmindedness between individuals living in one family does not depend upon the parents. However, for Radicalism, some substantial part of the environment may depend upon the parents. The size of e shows that a component independent of the parents may also be important. Although e is large for both traits, the greater p for Radicalism means that we are more likely to find cultural transmission as defined by Eaves (1976a) for Radicalism and it is this possibility to which we now turn. The model which we define in the next Section, considers only the environment dependent on the parent. However, we shall look at the importance of an independent environmental component in later Sections.

4 THE GENETICAL PATH MODEL

4.1 The Path Diagram

Eaves' (1976b) model shows how the transmission of cultural differences mediated through the phenotype of the parents leads to genotype-environmental covariance. In order to fit his model to our data, we would have to assume no assortative mating. The observed significant marital correlation in our data, provides direct evidence for assortative mating. Therefore, it is unrealistic to use Eaves' model without reformulating it to include assortative mating. Since this would be difficult, the simplest way to examine genotype-environment covariation and assortative mating simultaneously is to use the technique of specifying effects of interest using the path model approach. The full model including cultural transmission and assortative mating will be described first. Certain simplifying assumptions to be tested will be discussed later as they arise.

The full model yields a very complex path diagram and it has been split into two parts for tracing the paths. These parts are given in figures B6 and B7. These figures can be recombined as shown in Figure B8 to reproduce the original path diagram. Intergenerational relationships may be specified using Figure B6 and collateral relationships using Figure B7. Occasionally we may have to refer to Figure B8 but we shall avoid this as much as possible.

FIGURE 80: PARENT-OFFSPRING RELATIONSHIPS OF A GENETICAL PATH MODEL

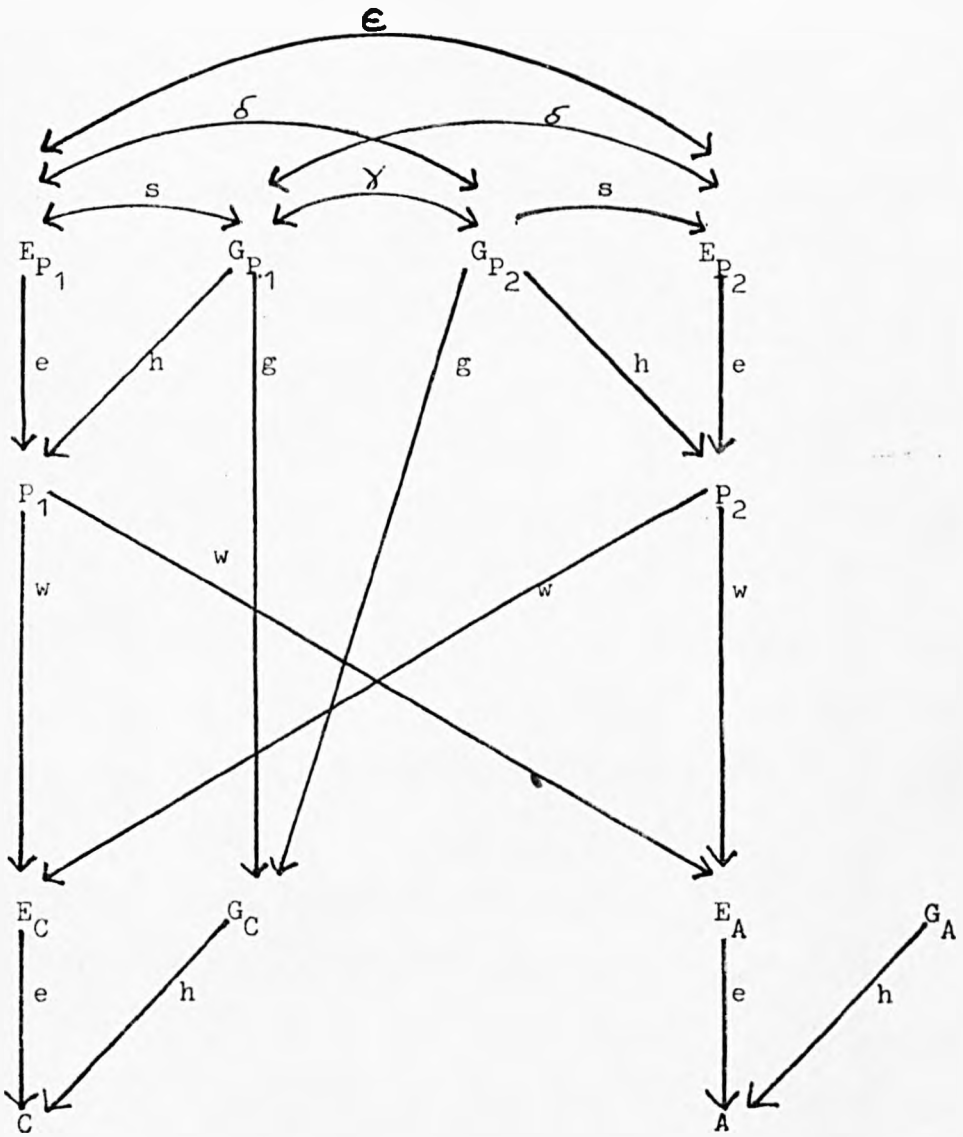


FIGURE B7: COLLATERAL RELATIONSHIPS FOR A GENETICAL PATH MODEL

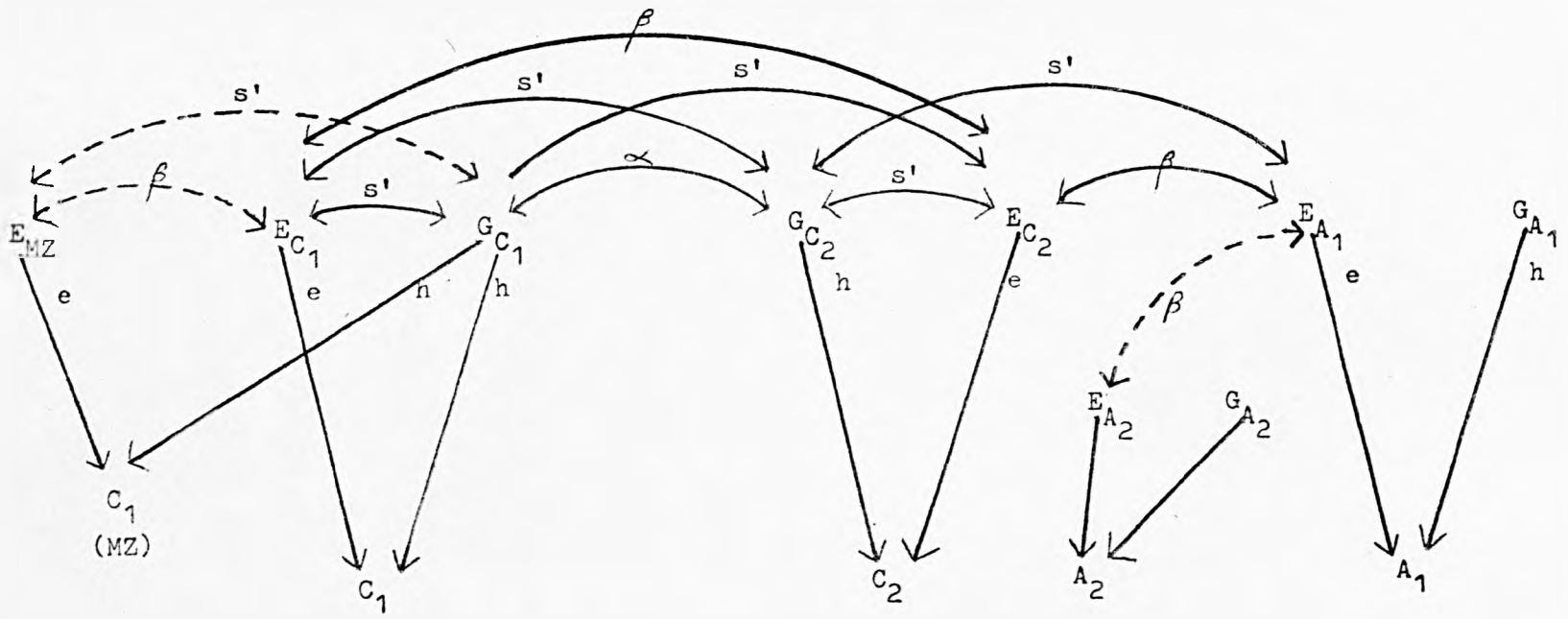
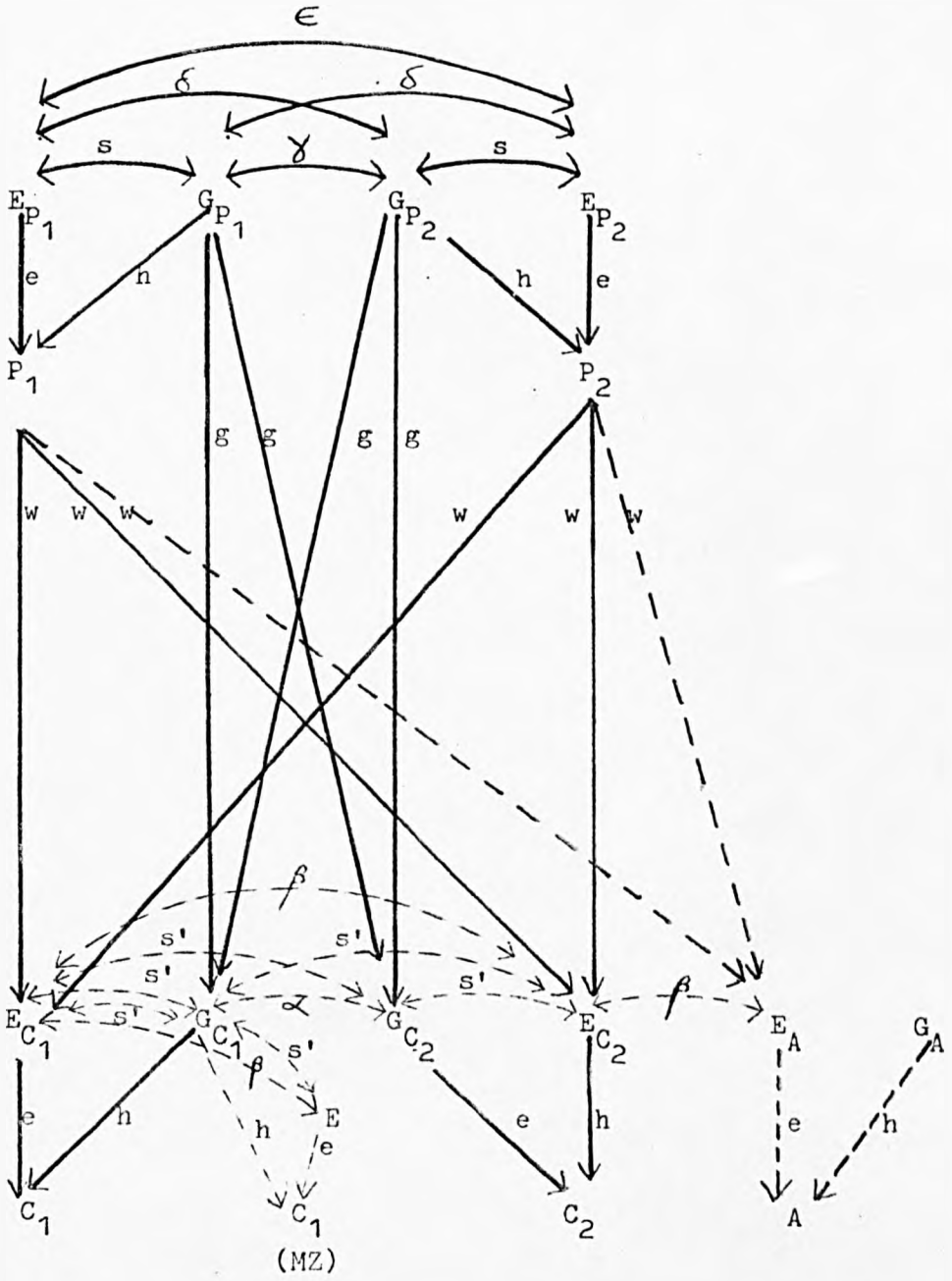


FIGURE B8: A GENETICAL PATH MODEL



4.2 Notation

The main types of individual are specified:

P - Parent

C - Child

A - Adopted Child

Child may sometimes include adopted child, when it is irrelevant whether a child is natural or not i.e. when we are considering environmental effects. C was used to include all types of children in the last section when fitting environmental path models.

Subscripts to P, C and A are used to indicate different individuals, where several individuals of the same type occur in one family e.g. P_1 and P_2 for first and second parent (using mother and father respectively). A special case occurs when we use C_1 twice in order to represent a pair of monozygotic twins.

The other capitals have the following meaning:

G is the genotype (an unknown)

E is the environment (an unknown)

Subscripts to G and E indicate the individual to who the G and E refer.

Paths are always denoted by lower case letters either Roman or Greek.

From Figure B7, α , β and s' are correlations among genotypes and environments implied by the relationships in the preceding generation, and can be expressed in terms of the paths in Figure B6.

Similarly, considering Figure B6, γ , δ , ϵ and s are implied correlations between the genes and environments in the parental generation which could be resolved by reference to the grand-parental generation.

In the collaterals, the correlation between any genotype and environment is given by s'. However in the parental generation genotype-environment correlations are given by s and δ . This is because the collaterals share a common environment provided by their parents and the correlations between G_{C_1} and E_{C_1} and between G_{C_1} and E_{C_2} can be regarded as identical. However in the parental generation, there is no reason why E_{P_1} should be the same as E_{P_2}, since the grandparents providing the environment of the parents are different, and, therefore, separate correlations are needed for the correlation between an individuals own genotype with his environment and the correlation between an individuals own genotype and his spouses environment. The correlation δ arises because each spouse selects the other on the basis of the phenotype and thus both genetical and environment correlations exist between them. The correlation s arises as a result of the transmission of genes and environment from the preceding generation, and will be discussed in more depth later.

The implied correlations represent the following:

s' - correlation between the genes of one sibling and either his own environment or that of his siblings

s - correlation between a parents genotype and environment

δ - correlation between a parent's genotype and his spouses environment

α - correlation between the genes of natural siblings, which arises because they have the same parents

γ - correlation between the genes of spouses, which arises if there if assortative mating

β - correlation between the environments of siblings (natural or adopted). The common environmental covariance is given by $e^2\beta$.

ϵ - correlation between the environments of spouses which arises if there is assortative mating.

The genotype-environment correlation terms can only appear, unless we totally redefine the model, when there is a path from the parents phenotype to the offspring's environment i.e. when there is cultural transmission of the type defined by Eaves (1976b). Eaves (1976a) has shown that genotype-environment covariance can also arise when there is co-operation or competition between siblings i.e. when there is a path from one siblings phenotype to anothers environment. This possibility is not considered in this study and would involve a totally different path diagram. It could also be studied using conventional biometrical-genetical techniques. Another source of genotype-environment-covariance arises when individual's select their own environment. This is not amenable to any sort of analysis since the covariance in this case is always confounded with the individual's genotype in non-experimental populations.

For the purposes of this study, therefore, s, s' and δ cannot appear unless there is a path from parental phenotype to offspring's environment. This path is called w.

The other paths are:

e - path from the environment to the phenotype

g - direct path from parent's genotype to offspring's genotype

h - path from genotype to phenotype

If gene action is purely additive, we expect that $g = \frac{1}{2}$. Therefore,

if we allow g to be a free parameter in the model, and its value differs significantly from $g = \frac{1}{2}$, we are forced to conclude that our model is inappropriate. This may be because of non-additive effects such as dominance or epistasis which cannot be specified using path diagrams, or other inadequacies in the model. The direction of departure from $g = \frac{1}{2}$ depends on such factors as the sample and will not, in general, be systematically related to such effects as dominance.

If there is no genotype-environment covariance, then h^2 is a direct estimate of the proportion of the total variation due to genetical differences. In our full model, h^2 is not an estimate of the heritability since correlations between genes and environment have been specified. However, this will not be the case when we consider several simplifying assumption later.

The model described in this section makes several assumptions. These are:

i No selective placement of the adopteds

ii No environmental contribution of the true parents to the adopted children

iii Random distribution of adopting parents throughout the population e.g. they are not concentrated in the higher socio-economic status groups

iv Maternal and paternal environmental contributions are equal.

4.3 Defining Expectations for the Variances

The model described in the previous sections specifies the effects of cultural transmission based on the phenotype, as described by Eaves (1976b). Such cultural transmission leads to genotype-environment

covariation which is designated s in the parental generation and s' in the offspring generation. If the population is in cultural equilibrium, we expect that s' = s. However, we have no reason to suppose that such an equilibrium has been reached for social attitudes, so s and s' will both be treated as free parameters in our model, in the first instance. Later we shall test the hypothesis that the population is in equilibrium by comparing the likelihood when the population is not in equilibrium i.e. $s' \neq s$ with that when we specify that the population is in equilibrium by constraining that s' = s.

When s' does not equal s, the total variances of parent's and offspring will differ. This means that, for the first time, we shall allow the diagonal terms of the expected covariance matrix to differ, since $V_P \neq V_C$. Cultural transmission also leads to inequality in the total variances of natural and adopted children. Eaves (1976b) has shown that when his coefficient of cultural transmission, b, is positive the total variance of individuals reared by their own parents, V_C, is greater than that of individuals reared by adopting parents, V_A. When b is negative, then $V_C < V_A$. Thus with the non-equilibrium cultural transmission model depicted in Figures B6, B7 and B8, $V_P \neq V_C \neq V_A$. i.e. we can have three different expected variances on tne diagonal of the expected covariance matrix.

The variance of adopteds, using standard path notation is given by:

$$R^2 = h\, r_{AG} + e r_{AE}$$

$$\therefore \quad R^2 = h^2 + e^2$$

Since $R^2 = 1$, we have:

$$1 = h^2 + e^2$$

i.e. the total variance has been standardised to unity. In order to work with variances, we multiply this equation by the total variance of the adopteds, to obtain:

$$V_A = V_A (h^2 + e^2).$$

This result is a tautology for the case of the adopteds, since $h^2 + e^2 = 1$. However, it proves useful in defining V_C and V_P.

We now define the unit variance of children reared by their natural parents as:

$$R^2 = hr_{CG} + er_{CE}$$

$$\therefore \quad R^2 = h(h + es') + e(e + s'h)$$

$$\text{i.e.} \quad R^2 = h^2 + 2ehs' + e^2$$

Therefore, as the non-standardised total variance we obtain:

$$V_C = V_A (h^2 + 2ehs' + e^2)$$

$$\longrightarrow V_C = V_A (1 + 2ehs')$$

i.e. we obtain the same result as Eaves (1976b) since when s' is positive, $V_C > V_A$ and vice versa.

We may similarly define V_P:

$$V_P = V_A(1 + 2ehs)$$

We can see, therefore, that when the population is in cultural equilibrium that $V_P = V_C \neq V_A$.

4.4 Defining Expectations for the Covariances

In Figure B7, the two individuals marked C_1 are identical twins, since they share a common G. Following the usual rules of path analysis,

(Li, 1955), we can show from this diagram that the covariance of MZ twins is given by:

$$COV_{MZ} = V_C (h^2 + 2ehs' + e^2\beta)$$

Similarly if C_1 and C_2 are either siblings or dizygotic twins, the covariance is given by:

$$COV_{DZ} = COV_{SIB} = V_C (h^2\alpha + 2ehs' + e^2\beta)$$

If we look at the difference between MZ and DZ twins,

$$COV_{MZ} - COV_{DZ} = V_C h^2 (1 - \alpha)$$

which is equivalent to G_1 in conventional biometrical genetical terminology i.e. the variation within families produced by the segregation of the genes.

The covariance of the two adopted individuals, A_1 and A_2 may be expressed:

$$COV_{AA} = V_A (e^2\beta)$$

This differs from the covariance of adopted and natural siblings reared together, because the total variance of the two groups differs. This is given by:

$$COV_{AC} = \sqrt{V_A \times V_C}\ (e^2\beta)$$

It would be possible to define the marital correlation:

$$\mu = h^2\gamma + 2eh\delta + e^2\epsilon$$

However, this would mean that we could not resolve γ, δ, ϵ and s in terms of the other parameters and we shall, therefore, estimate the marital correlation directly from the data by specifying the covariance between spouses:

$$COV_{P_1P_2} = V_P \mu$$

The expectation for the natural parent-offspring covariance is complex. We derive this from figure B6. In order that the reader may follow the derivation of this expectation, we will write out the paths in full in the order in which they are traced. This gives:

$$COV_{PC} = \sqrt{V_P \times V_C}\ (hgh + hgse + hg\gamma h + hg\delta e$$

$$+ ew + ewh\delta h + ewh\delta e + ewe\delta h + ewe\epsilon e)$$

which simplifies to:

$$COV_{PC} = \sqrt{V_P \times V_C}\ (gh^2(1+\gamma) + egh(s+\delta) + ew(1+h^2\gamma + 2eh\delta + e^2\epsilon))$$

The adopting parent-offspring covariance may be similarly traced from Figure B6 and turns out to be:

$$COV_{PA} = \sqrt{V_P \times V_A}\ (ew(1 + h^2\gamma + 2eh\delta + e^2\epsilon))$$

Thus, we now have expectations for all possible elements of the expected covariance matrix. These involve fifteen parameters. We could not hope to estimate this many parameters from the data. However, these parameters are not all free parameters and we can resolve some of them in terms of the others to give us a final model of eight parameters: μ_m, μ_f, V_A, g, h, w, s and μ.

Thus, from Figure B8, we can show that:

$$\alpha = 2g^2(1 + \gamma)$$

$$\beta = 2w^2\ (1 + 2eh\delta + e^2\epsilon + h^2\gamma)$$

$$s' = 2gw\ (es + h\delta + h + h\gamma)$$

which eliminates three of the original fifteen parameters. Similarly from Figure B6, we can show that:

$$\gamma = \mu (h + es)^2$$

$$\delta = \mu (hs + e)(h + es)$$

$$\epsilon = \mu (e + hs)^2$$

Also, we know that $e = \sqrt{1 - h^2}$. Therefore, making these substitutions into the expectations for the variances and covariances, we arrive at a final non-equilibrium model of 8 parameters. These expectations are summarised in Table 50, in terms of all the paths and implied correlations. The substitutions that can be made in order to simplify the number of parameters to eight are given below the expectations. These substitutions were never made into the expectations by hand, but were specified in the computer program.

4.5 Fitting the Full Eight Parameter Path Model

This model, described above, includes genetical and environmental effects, genotype-environment covariation produced by cultural transmission and assortative mating. In order to fit this model, we set up the data files in the usual way, as described in Section 7. The model is specified in the program as described earlier. In order to specify estimates of minus the maximum log likelihood, we examined previous analyses. Since the estimates must be less than the final solutions for minus the log likelihood, based on the present model, we chose fairly low values of 500.0 and -1350.0 for Radicalism and Toughmindedness respectively.

The only remaining problem is to choose sensible trial values. The values calculated from the data in Section 5 were given for μ_m,

TABLE B 50 : EXPECTATIONS FOR THE TERMS OF THE COVARIANCE MATRIX

VARIANCE OR COVARIANCE OF :	EXPECTATION
ADOPTED CHILDREN (V_A)	$V_A (h^2 + e^2)$
NATURAL CHILDREN (V_C)	$V_A (h^2 + 2ehs^1 + e^2)$
PARENTS (V_p)	$V_A (h^2 + 2ehs + e^2)$
MZ TWINS	$V_C (h^2 + 2ehs^1 + e^2\beta)$
DZ TWINS OR SIBLINGS	$V_C (h^2\propto + 2ehs^1 + e^2\beta)$
NATURAL AND ADOPTED CHILDREN	$\sqrt{V_A \times V_C}\,(e^2\beta)$
TWO ADOPTED CHILDREN	$V_A\ (e^2\beta)$
NATURAL PARENT CHILD	$\sqrt{V_p \times V_C}\ (gh^2(1+\gamma)+egh(s+\delta)+ew (1+h^2\gamma+2eh\delta+e^2\epsilon))$
ADOPTING PARENT CHILD	$\sqrt{V_p \times V_A}(ew\ (1+h^2\gamma+2eh\delta+e^2\epsilon))$
SPOUSES	$V_p\,\mu$

The following substitutions can be made into the above expectations:

$e = \sqrt{1-h^2}$

$\gamma = \mu(h+es)^2$

$\delta = \mu(hs+e)\ (h+es)$

$\epsilon = \mu(e+hs)^2$

$\propto = 2g^2\ (1+\gamma)$

$\beta = 2w^2\ (1+2eh\delta+e^2\epsilon+h^2\gamma)$

$s^1 = 2gw\ (es+h(1+\gamma+\delta))$

μ_f, V_A, and μ, and trial values of 0.5 were given for the paths g, h, w and s. Convergence took some time, but was achieved using the method based on first derivatives and no constraints were necessary. Parameter estimates for Radicalism and Toughmindedness and their corresponding maximum log likelihoods are presented in Table B51. Values for the variances, implied correlations and paths which may be specified in terms of the parameters of the model were calculated and given in the lower part of Table B51. We will not consider the values of the parameters of this model in any detail since we want to move on and compare this model with a similar but simplified model. However, we shall briefly compare the maximum log likelihoods with those found for earlier models. For both traits, the maximum log likelihoods based on this model are greater than those for any of our previous models. In the case of Radicalism, this model does not provide a significant improvement in fit over the $E_1D_RE_2'$ model ($\chi^2_3 = 1.5972$, $50\% < p < 70\%$), but it does enable us to partition the between families variation into that due to assortative mating and that due to environmental differences dependent upon parental phenotype.

This genetical path model provides a significant improvement in fit over the most likely model we have found for Toughmindedness. This was the environmental path model 2B whose maximum log likelihood was 1314.7409. Thus, using the likelihood ratio criterion, $\chi^2_2 = 28.8022$, $p < 0.1\%$. The main difference between the two models is that in the latter model, there is genetical variation and genotype-environmental covariance leading to inequality of the total variances of parents, natural children and adopted children. In Section 5, we saw differences in total variance between these three groups. However, there was no

TABLE B 51 : FULL EIGHT PARAMETER GENETICAL PATH MODEL

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0172	0.0093
μ_f	-0.0042	0.0065
V_A	0.8489	0.1010
g	0.5477	0.5652
h	0.8489	0.8988
w	-0.0539	-0.2106
s	0.0156	-0.1103
μ	0.3173	0.4692
L	-509.0304	1329.1420
e	0.5286	0.4386
γ	0.2331	0.3393
δ	0.1474	0.1355
ϵ	0.0932	0.0541
α	0.7398	0.8557
β	0.0077	0.1234
S^1	-0.0697	-0.3040
V_c	0.7958	0.0768
V_p	0.8608	0.0922

evidence for any genetical variation in the twin data. This model may provide a significant improvement over previous models because it predicts differences in the total variances or it may be the first indication that we have for genetical variation in Toughmindedness. We will consider the implications of this possibility and why there was no evidence for it in the twin data later.

4.6 A Seven Parameter Genetical Path Model

In the previous model, eight parameters were fitted which included g, the path from parental genotype to offspring genotype. The value of this path should be 0.5 if all the variation is additive. But if g differs significantly from 0.5, we may take this as evidence for some form of non-additivity in the data. We see that $g = 0.5477$ for Radicalism and $g = 0.5652$ for Toughmindedness. These values are close to 0.5, but we need a test of significance to determine whether we can regard them as estimates of the genetical path which differ from 0.5 only by sampling error. In order to make this test of significance we shall fit a simplified model to the data.

This model is similar to that specified in Table B50, but 0.5 is substituted for g in all the expectations. This yields a 7 parameter genetical path model: μ_m, μ_f, V_A, h, w, s and μ. In order to obtain the maximum likelihood solutions for these parameters, we reran the computer program using the new expectations (with $g = \frac{1}{2}$), the same estimates of minus the log likelihood and the solutions for the parameters from Table B51 as trial values. The method based on first derivatives was used for both traits. Convergence was reasonably quickly reached

for Radicalism. However, in the case of Toughmindedness the program failed after a number of iterations, with the covariance matrix becoming non-positive definite. The matrix was constrained to be positive definite by specifying that all its eigenvalues must be greater than zero and the program was rerun. This time convergence was quickly achieved. The parameter estimates and the maximum log likelihoods are given in Table B52 for the unconstrained solution for Radicalism and the constrained solution for Toughmindedness.

Now we will compare the likelihoods for $g = \frac{1}{2}$ and $g \neq \frac{1}{2}$. Using the likelihood ratio criterion, the eight parameter model for Radicalism does not provide a significant improvement over the seven parameter model. ($\chi^2_1 = 1.0618$, $30\% < p < 50\%$). Therefore, we have no reason to suppose that g takes any other value than 0.5 i.e. there is no evidence for genetical non-additivity. This is consistent with our analysis of the twin study.

However, in the case of Toughmindedness, we find a different picture. The likelihood ratio criterion gives us a value of chisquare: $\chi^2_1 = 4.0894$, $2\% < p < 5\%$, which is significant at the five per cent level. We can attribute this significant difference between the seven and eight parameter models to the fact that $g \neq \frac{1}{2}$. This indicates the inadequacy of the additive model for these data.

4.7 Discussion of Toughmindedness

In Section 6, we found no evidence for any genetical component of Toughmindedness, as measured by the Wilson-Patterson Attitudes

TABLE B 52 : SEVEN PARAMETER MODEL, CONSTRAINING THAT ALL GENETICAL VARIATION IS ADDITIVE

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0221	0.0104
μ_f	-0.0048	0.0069
V_A	0.8462	0.0998
h	0.8573	0.8976
w	-0.0543	-0.1704
s	0.0655	0.0422
μ	0.4306	0.5802
L	-509.5613	1327.0973
ε	0.5000	0.5000
e	0.5148	0.4408
γ	0.3419	0.4870
δ	0.2191	0.2545
ϵ	0.1404	0.1329
α	0.6710	0.7435
β	0.0087	0.0941
s^1	-0.0745	-0.2695
V_c	0.7905	0.0785
V_p	0.8951	0.1031

Questionnaire and concluded that variation could be explained by unreliability of measurement, specific environmental effects and a common environmental component. Yet now we find that the genetical path model provides a significant improvement over environmental models. It is difficult to understand this discrepancy, but two explanations are possible.

The first is that the great improvement of this model over previous models is purely a result of relaxing the constraint that the total variances are equal. In Section 5, we saw that $V_P = 0.104 > V_A = 0.100 > V_C = 0.075$. This model allow the variances to differ in this order given appropriate values of s and s'.

If this is not the reason for the genetical path model providing the best explanation of the data, then there may be genetical variation determining individual differences in Toughmindedness. Why then did we not detect genetical variation in the twin data? Perhaps non-additive variation and interactions could produce the sort of results we have found, if they were confounded with estimates of E_1 and E_2 in the twins, but not in the adoption data. Since $g \neq \frac{1}{2}$, we know that the genetical path model for Toughmindedness is not appropriate. Non-addivity of environmental effects or sex interactions, without a genetical background, might provide the most appropriate explanation for our findings. Our tentative conclusion is that there may be some genetical variation for Toughmindedness of which a large proportion is non-additive. (We recall that $\chi^2_1 = 3.71$, $5\% < p < 10\%$, when we compared E_1BD_R and E_1E_2 models in the twins, which approaches significance and that all three parameters of the E_1BD_R model were significant. Thus, genetical variation was not strongly precluded in the twins). However, it is also possible that

the variation may be explained by unequal environmental effects in the different groups, which lead to inequality of the total variances; non-additive environmental effects, or sex interactions.

A comparison of the means, total variances and marital correlation calculated from the data and calculated from the estimates of the seven parameter and eight parameter genetical path models is given in Table B53. Estimates of the means and total variances are similar in all cases, giving us no additional information. Estimates of the marital correlation based on the two models are rather high. However, the estimate of μ based on the eight parameter model is closer to the true value than that based on the seven parameter model, providing additional evidence that the eight parameter model is more appropriate for our data.

In both models w is negative. However, s and s' take the same sign in the eight parameter model whereas in the seven parameter model, only s' is negative. A mechanism whereby the covariance would differ in direction between the generations is hard to visualise, and the eight parameter model in which s and s' are both negative again seems more reasonable. Considering now only the eight parameter model, the genotype-environment correlations are quite large and s' is substantially bigger than s. i.e. the negative covariance between genotype and environment has increased between the two generations. This suggests that the population is not in equilibrium. The concept of cultural equilibrium, its algebraic definition and the solution of an equilibrium model will be discussed later in reference to Radicalism. However, the simplest way to achieve an equilibrium solution for the parameters of the path model using the computer program is simply to specify the constraint that $s = s'$. This was done for the seven and eight

TABLE B 53: COMPARISON OF ESTIMATES FROM THREE SOURCES FOR TOUGHMINDEDNESS

	ESTIMATES FROM SECTION	ESTIMATES FROM SEVEN PARAMETER MODEL	ESTIMATES FROM EIGHT PARAMETER MODEL
μ_m	0.010	0.010	0.009
μ_f	0.001	0.007	0.007
V_A	0.100	0.100	0.101
V_C	0.075	0.079	0.077
V_p	0.104	0.103	0.092
μ	0.382	0.580	0.469

parameter models. The values of the unconstrained solution were given as trial values. Solutions were sought using function values only. he m thod using first derivatives and the method using second derivatives. In all cases the covariance matrix became non-positive definite and the programs failed. They were rerun with the additional constraints needed to ensure that the covariance matrix remained positive definite. These programs ran for some time and were restarted using the current solutions as new trial values. However, movement towards convergence was very slow and was obviously not going to be reached in a reasonable amount of time. Thus solutions for the parameters of the equilibrium solution were never found. The great difficulty encountered suggests that the equilibrium model may be inappropriate and that the population is not in equilibrium. This is reasonable in view of the large difference between the covariance terms of the two generations.

Thus, the eight parameter, non-equilibrium genetical path model remains our most likely model for Toughmindedness. Values of the implied correlations and the paths all seem reasonable and lie within the expected range. However, we will defer further discussion of Toughmindedness until we have considered the likelihoods of certain simplified models.

4.8 Discussion of Radicalism

Since we have considerable genetical variation for Radicalism, but have found no evidence to suggest non-additivity, the seven parameter genetical path model is more appropriate than the eight parameter model. Although this model is not significantly more likely

than the $E_1D_RE_2'$ model, it provides a more useful parameterisation, which enables us to consider certain effects of interest.

A comparison of estimates of the means, variances and marital correlation calculated from the data in Section 5 and calculated from the two genetical path models is given in Table B54. Estimates of the means and total variances are similar and the total variances take the same order i.e. $V_P > V_A > V_C$. However, we have no evidence that lifting the constraint that the total variances must be equal significantly improves the likelihood, when we compare the likelihoods based on this model and that based on the $E_1D_RE_2'$ model. Again, estimates of the marital correlation differ in the three cases, and the estimate from the most likely model is closest to the true value, providing further evidence that the seven parameter model is the most appropriate.

Estimates of the paths and implied correlations seem reasonable and internally consistent. However, we will defer discussion of the relative magnitudes of the parameters and their interpretation until we have considered the equilibrium solution.

4.9 Equilibrium Solution for Radicalism

The path from parent's phenotype to offspring's genotype is negative suggesting that the environmental effect of more radical parents is to make their children less radical and vice versa. The effect is only small, as judged by the size of w. The genotype-environment correlation is also small but differs between the two generations by a small amount.

TABLE B 54 : COMPARISON OF ESTIMATES FROM THREE SOURCES FOR RADICALISM

	ESTIMATES FROM SECTION 5	ESTIMATES FROM SEVEN PARAMETER MODEL	ESTIMATES FROM EIGHT PARAMETER MODELS
μ_m	0.005	-0.022	-0.017
μ_f	-0.002	-0.005	-0.004
V_A	0.845	0.846	0.849
V_C	0.762	0.791	0.796
V_p	1.011	0.895	0.861
μ	0.405	0.431	0.317

If the population is in equilibrium we expect that s and s' will be equal. We would like to know whether s and s' differ significantly i.e. whether population is in equilibrium. In order to do this we need to fit a model assuming equilibrium and compare the maximum log likelihoods of equilibrium and non-equilibrium solutions.

In order to assume equilibrium we let s' = s. In Table B50, we saw that s' may be expressed in terms of s, as follows

$$s' = 2gw\ (es + h(1 + \gamma + \delta\ )$$

Now at equilibrium s' = s. Therefore, substituting and transferring terms in s to the left hand side we find:

$$s(1 - 2egw) = 2gwh(1 + \gamma + \delta\ ) \qquad \ldots 1$$

which simplifies to:

$$s = \frac{2ghw(1 + \gamma + \delta\ )}{1 - 2egw}$$

However, from Table B50, we see that γ and δ contain terms in s. Therefore, substituting in equation 1, we obtain:

$$s(1-2egw) = 2ghw(1+ \mu h(e+h)+ s\mu(e+h)^2 + s^2 \mu e(e+h))$$

which we may express:

$$s^2(2eghw\mu(e+h))+ s(\mu(e+h)^2- (1-2egw)) + 2ghw(1+ \mu h(e+h)) = 0$$

This quadratic in s may be expressed as:

$$As^2 + Bs + C = 0, \quad \text{where}$$

$$A = 2eghw\mu(e+h)$$

$$B = \mu(e+h)^2 - (1-2egw)$$

$$C = 2ghw\ (1+ \mu h(e+h))$$

Therefore, s is given by:

$$s = -\frac{B \pm \sqrt{B^2 - 4AC}}{2A} \qquad \ldots 2$$

This equation has two roots and thus two solutions of s are possible which produce an equilibrium. Clearly, in practise only one of these solutions is likely to be sensible. However, we have no way of knowing which this will be, and it may turn out not to be the same root in all situations.

In order to fit the equilibrium path model using our expression for s, defined by equation 2, we would have to reformulate the expectations in Table B50. Moreover, we would have to set up two sets of expectations based on the two possible roots for s, and attempt to find solutions for each of these sets of expectations, which would be expensive in computer time. It might not even be possible to achieve a solution in this way, because oscillation might occur between two minima. However, we can easily overcome these problems by constructing a penality function, added to L, which tends to zero when $(s' - s) = 0$. In order to obtain solutions for the equilibrium model of the seven and eight parameter genetical models, we used the estimates from the non-equilibrium model as trial values. The method based on first derivatives was used to obtain the maximum log likelihood solution for the parameters of the model. Convergence was rapidly reached without the need to constrain the covariance matrix to be positive definite. The ease with which a solution was found for Radicalism suggests that perhaps the equilibrium model is appropriate, whereas for Toughmindedness it was clearly not appropriate and an equilibrium solution could not be found.

Estimates of the parameters for eight and seven parameter-models and their likelihoods are given in Tables B55 and B56 respectively. The paths and implied correlations which can be resolved in terms of these parameters were calculated and are given in the bottom part of these tables. Comparison of the maximum log likelihoods of these models with those of the non-equilibrium models will show whether the non-equilibrium solution is more likely than the equilibrium solution. Since there is one less free parameter in the equilibrium model, we can use the likelihood ratio criterion to make a test of significance. For the eight parameter model $\chi^2_1 = 0.8926$, $30\% < p < 50\%$, and for the seven parameter model $\chi^2_1 = 2.6202$, $10\% < p < 20\%$. Whichever model we consider, specifying the constraint that $s' = s$ does not produce a significant reduction in the likelihood. Thus there is no evidence that the genotype-environment correlation differs in the two generations being considered.

Comparing the eight and seven parameter model using the likelihood ratio criterion, we find that $\chi^2_1 = 2.6894$, $10\% < p < 20\%$. Therefore, we have no reason to suppose that g differs significantly from 0.5. The most appropriate path model for our data so far is the model whose solutions are given in Table B56. In this model, there is additive genetical variation, assortative mating and cultural transmission as defined by Eaves (1976b), with the population in cultural equilibrium. The value of h is large suggesting that there is a large heritable component of variation. Following on from this we see that there is a large genetical correlation between siblings, α. The assortative mating leads to a substantial correlation between the genes of spouses i.e. $\gamma = 0.2224$, but a much lower correlation between their

TABLE B 55 : EQUILIBRIUM SOLUTION FOR FULL EIGHT PARAMETER MODEL

PARAMETER	RADICALISM
μ_m	-0.0176
μ_f	-0.0043
V_A	0.8315
ε	0.5473
h	0.8311
w	-0.0234
s	-0.0305
μ	0.3152
ML	-509.4767
e	0.5561
γ	0.2089
δ	0.1362
$\in$	0.0888
α	0.7242
β	0.0014
s^1	0.0282
V_C	0.8532
V_p	0.8080

TABLE B 56 : EQUILIBRIUM SOLUTION FOR SEVEN PARAMETER MODEL

PARAMETER	RADICALISM
μ_m	-0.0190
μ_f	-0.0043
V_A	0.8302
h	0.8405
w	-0.0257
s	-0.0290
μ	0.3270
ML	-510.8214
g	0.5000
e	0.5418
γ	0.2224
δ	0.1396
ϵ	0.0875
α	0.6112
β	0.0017
s^1	-0.0290
V_C	0.7828
V_p	0.8083

environments, $\epsilon = 0.0875$. The correlation between the environments of siblings provided by their parents is small: $\beta = 0.0017$, suggesting that any common environmental component dependent upon the parents is quite small. This is a consequence of the low value of the path, w, from parental phenotype to offspring's environment, which suggests that cultural transmission may be making only a minimal contribution to the total variation. The small genotype-environment correlation which arises as a result of cultural transmission is estimated to be small suggesting that any effect of cultural transmission is minimal. We, therefore, need to test whether the effect of cultural transmission is significant by comparing the likelihood of this model with that based on a simplified model in which there is no cultural transmission.

5 SIMPLIFIED GENETICAL PATH MODELS

5.1 Introduction

We know from the twin data that there is evidence for a between families component produced by either common environmental influences or by assortative mating. This was confirmed by fitting biometrical-genetical models to all the data jointly using the methods of pedigree analysis. However, we were still unable to discriminate properly between the two alternative sources of the between families component.

In the last section we considered the possibility that both assortative mating and cultural differences mediated through the phenotype of the parents were important. We found that the assortative mating parameters were substantial, but that the path, w, and the genotype-environment covariance were minimal for Radicalism. This suggested that there might be no significant variation produced by cultural transmission. How then can we interpret the between families component? There are two possibilities. All the between families component could be due to assortative mating. Alternatively there could be a common environmental effect independent of the parents. Or, both these effects might be contributing.

If we test the first possibility by specifying a path model with no cultural transmission, but with assortative mating, we can use this model to test the significance of the variation produced by cultural transmission, by comparing it with the model discussed at length in Section 8.4.

We can also specify a model with a common environmental component and determine whether this gives us a more likely explanation of the observed variation in attitudes.

5.2 Simplified Genetical Path Model

A path diagram for a genetical model with assortative mating but no common environmental component is given in Figure B9, by the solid lines. The model is similar to that used in Section 8.4, except that there is now no path from parental phenotype to offspring environment. Since we do not have this path, specifying cultural transmission, we no longer have any genotype-environment covariation i.e. there are no implied correlations: s, and s'. Considering the model given by the solid lines, there is no correlation between the environment of siblings i.e. there is no common environmental component. E now represents only specific variation. A correlation between the genes of siblings is produced by the genetic system. Assortative mating gives rise to correlations between the genes of spouses and between their environments, since it is based on the phenotype. However, there is no longer any genotype-environment covariance in the parental generation. In effect, we now have an additive genetical model, with assortative mating and specific environmental variation.

This model is much simpler than the full model described in Section 8.4 and derivation of expectations for the expected covariance matrix presents no problems. These expectations, worked out from the solid lines in Figure B6 are given in Table B57.

FIGURE B9: SIMPLIFIED GENETICAL PATH MODEL

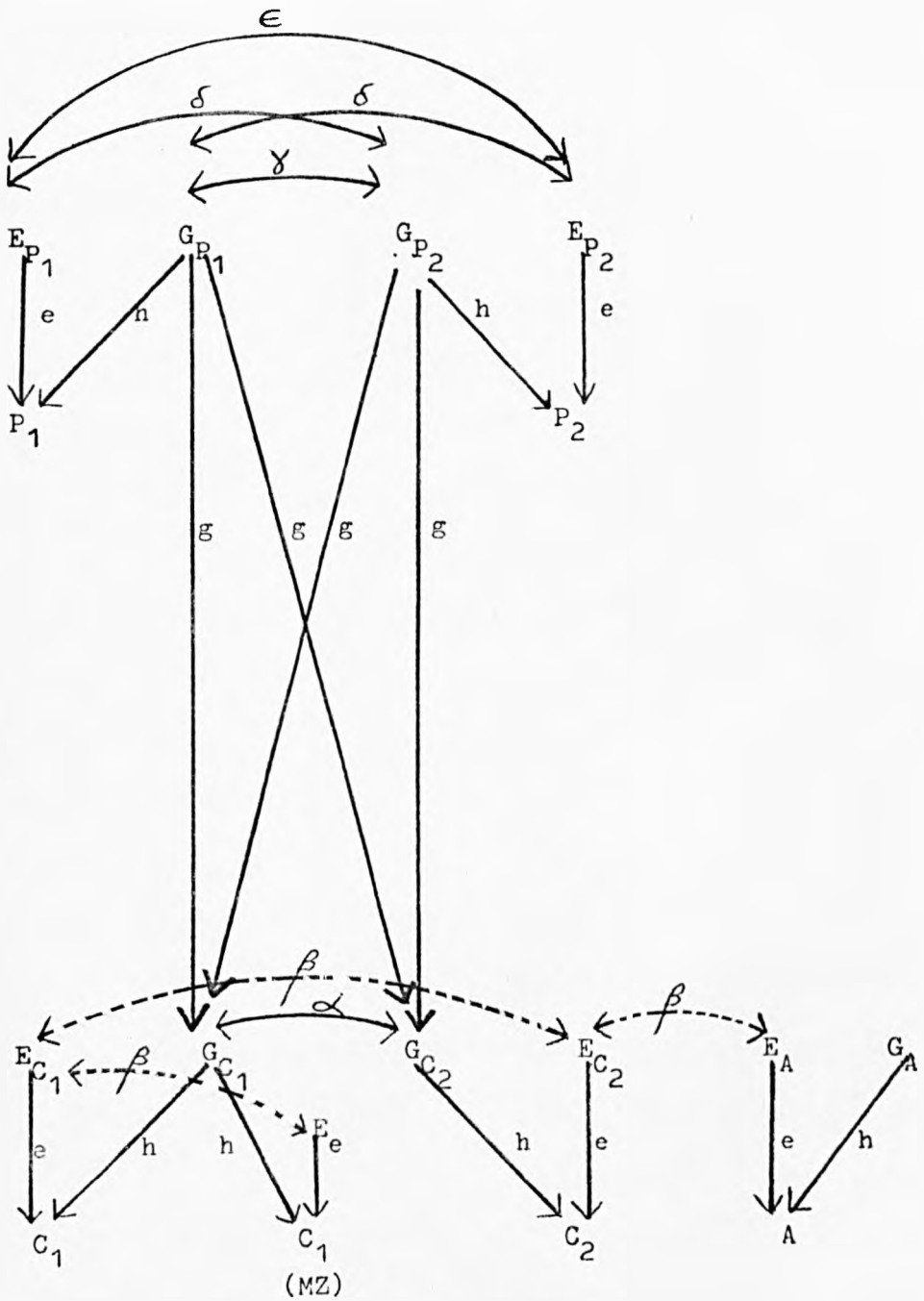


TABLE B 57 : EXPECTATIONS FOR A SIMPLIFIED GENETICAL MODEL

VARIANCE OR COVARIANCE	EXPECTATION
TOTAL VARIANCE	V
MZ TWINS	$V h^2$
DZ TWINS OR SIBLINGS	$V h^2 \alpha$
NATURAL AND ADOPTED CHILDREN	0.0
TWO ADOPTED CHILDREN	0.0
NATURAL PARENT-CHILD	$Vgh^2(1 + \gamma)$
ADOPTING PARENT-CHILD	0.0
SPOUSES	$V\mu$

The following substitutes can be made into the above expectations

$e = \sqrt{1 - h^2}$

$\gamma = \mu h^2$

$\epsilon = \mu e^2$

$\alpha = 2g^2 (1 + \gamma)$

We fitted this model to both Radicalism and Toughmindedness in order to determine whether it provides a more economical explanation of the data than the full genetical path model described in Section 8.4. The data file for the program was set up in the usual way. Two models were specified; one in which g was allowed to take its own value (6 parameters); the other where g was fixed at 0.5, by substituting 0.5 into the expectations given in Table B57 (i.e. 5 parameters). Values of 500.0 and -1350.0 were again given for the trial minus log likelihoods for Radicalism and Toughmindedness respectively. The solutions from the appropriate full genetical path model were given as trial values. The program was run using the method based on first derivatives. In each case the covariance matrix became non-positive definite and the program failed. However, when the program was rerun using the method based on second derivatives, convergence was rapidly reached for both models for Radicalism and Toughmindedness. The maximum likelihood solutions of the parameters and the log likelihoods are given in Tables B58 and B59 for the six and five parameter models respectively. Values of e and the implied correlations were calculated from the estimated parameters and are given below the likelihoods in these tables. We will defer discussion of these models until we have looked at a final model with a common environment independent of the parents, when we will compare the results for all the models fitted and determine which is the "best" model for each trait.

TABLE B 58 : SIMPLIFIED GENETICAL PATH MODEL WITH $g = \frac{1}{2}$

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0247	0.0111
μ_f	-0.0053	0.0099
V	0.8114	0.0833
g	0.5412	0.5672
h	0.8148	0.7860
μ	0.3400	0.4398
L	-509.1514	1324.7638
e	0.5845	0.6182
γ	0.2257	0.2717
$\in$	0.1162	0.1681
α	0.7180	0.8183

TABLE B 59 : SIMPLIFIED GENETICAL PATH MODEL WITH $g = \frac{1}{2}$

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0243	0.0116
μ_f	-0.0049	0.0097
V	0.8106	0.0831
h	0.8228	0.8001
μ	0.3893	0.5022
L	-510.0310	1322.7811
g	0.5000	0.5000
e	0.5683	0.5999
γ	0.2636	0.3215
ϵ	0.1257	0.1807
α	0.6318	0.6608
h^2	0.6770	0.6402
e^2	0.3230	0.3599

5.3 Simplified Genetical Path Model with an Independent Common Environment

We found some evidence from the environmental path models in Section 8.3 that the effect of an environment dependent on the parents was relatively small, whereas an environment independent of the parents was important in determining individual differences in attitudes. In order to test the importance of an independent common environment, we shall specify a model including its effect and compare its likelihood with that based on the model specified in the previous section. This model is given in Figure B9, if we consider both solid and broken lines. The model is similar to that described in the previous section, except that we now introduce a covariance between the environments of siblings reared together, which we call β. This changes our expectations for the covariances between collaterals. These expectations may be easily traced from the path diagram and are presented in Table B60.

In order to obtain the maximum likelihood solution for this model, we respecified the expected covariance matrix as shown in Table B60. The same trial values were used as for the previous model and a trial value of 0.2 was arbitrarily specified for β. Convergence was quickly reached in all cases using the method based on second derivatives. Solutions for the situations where $g \neq \frac{1}{2}$ and where $g = \frac{1}{2}$ are given in Tables B61 and B62 for Radicalism and Toughmindedness.

Having the results for all our models we are now in a position to determine the most appropriate model for the observed variation in all the data for each trait and to discuss the implications of this model.

TABLE B 60: EXPECTATIONS FOR SIMPLIFIED GENETICAL MODEL WITH COMMON ENVIRONMENT INDEPENDENT OF THE PARENTS.

VARIANCE OR COVARIANCE	EXPECTATION
TOTAL VARIANCE	V
MZ TWINS	$V\,(h^2 + e^2\beta)$
DZ TWINS OR SIBLINGS	$V\,(h^2\alpha + e^2\beta)$
NATURAL AND ADOPTED CHILDREN	$V\,e^2\beta$
TWO ADOPTED CHILDREN	$V\,e^2\beta$
NATURAL PARENT-CHILD	$V\,gh^2\,(1+\gamma)$
ADOPTED PARENT CHILD	0.0
SPOUSES	$V\mu$

The following substitutions can be made into the above expectations

$e = \sqrt{1 - h^2}$

$\gamma = \mu h^2$

$\epsilon = \mu e^2$

$\alpha = 2g^2\,(1+\gamma)$

TABLE B 61 : SEVEN PARAMETER MODEL WITH COMMON ENVIRONMENT INDEPENDENT OF THE PARENTS, $g \neq \frac{1}{2}$

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0237	0.0104
μ_f	-0.0055	0.0092
V	0.8117	0.0833
g	0.5327	0.5642
h	0.7603	0.6911
μ	0.3376	0.4295
β	0.2058	0.2715
L	-509.0011	1325.7574
e	0.6496	0.7228
γ	0.1952	0.2051
ϵ	0.1425	0.2244
α	0.6783	0.7672

TABLE B 62 : SIX PARAMETER MODEL WITH COMMON ENVIRONMENT INDEPENDENT OF THE PARENTS, $g = \frac{1}{2}$

PARAMETER	RADICALISM	TOUGHMINDEDNESS
μ_m	-0.0226	0.0106
μ_f	-0.0055	0.0087
V	0.8114	0.0830
h	0.7021	0.6428
μ	0.3462	0.4466
β	0.3482	0.3717
L	-509.2089	1324.8996
g	0.5000	0.5000
e	0.7121	0.7660
γ	0.1707	0.1845
ϵ	0.1756	0.2620
α	0.5854	0.5923

5.4 The Final Model for Radicalism

In order to decide which model of those we have fitted is the most appropriate for Radicalism we have prepared a summary of all the models fitted, which is given in Table B63. This Table gives the parameters of the models and their corresponding maximum log likelihoods.

The model with the highest maximum log likelihood is number 17, with L = -509.0011. This model, specifying the effects of genes, assortative mating and a common environment independent of the parents, will be used as a baseline against which to compare all the other models. We see immediately that models based purely on environmental variation between individuals (i.e. models 1, 2, 7, 8, 9, 10) are all significantly less likely than model 17. Therefore, our first conclusion, which confirms the results of the twin study, is that genetical differences determine differences between individuals in their Radicalism scores.

These environmental models, although not the most likely explanations of the data do contribute useful information. A comparison of the simple correlation models (1 and 2) shows that monozygotic twins are significantly more alike than dizygotic twins, suggesting either a genetical component of variation or that MZ twins are treated more alike. The importance of assortative mating is indicated by the comparison of model 7 with 8 and of 9 with 10. In both cases, the addition of the assortative mating parameter produces a significant difference in the likelihood, as measured by the likelihood ratio criterion. Using this criterion, we found that models including an environment independent of

TABLE B 63: SUMMARY OF MODELS FOR RADICALISM

MODEL	PARAMETERS	LIKELIHOOD	
1	μ_m μ_f V r	-533.7704	Simple correlation models
2	μ_m μ_f V r t	-515.8289	
3	μ_m μ_f E_1 D_R	-517.9692	Biometrical-genetical models
4	μ_m μ_f E_1 D_R E_2	-513.3635	
5	μ_m μ_f E_1 D_R E_2^{1}	-509.8290	
6	μ_m μ_f E_1 D_R E_2^{11}	-514.3745	
7	μ_m μ_f V p	-553.7629	Environmental path models
8	μ_m μ_f V p μ	-547.4797	
9	μ_m μ_f V p e	-523.4731	
10	μ_m μ_f V p e μ	-519.4698	
11	μ_m μ_f V_A g h w s μ	-509.0304	Full genetical path models
12	μ_m μ_f V_A h w s μ	-509.5613	
13*	μ_m μ_f V_A g h w s μ	-509.4767	
14*	μ_m μ_f V_A h w s μ	-510.8214	
15	μ_m μ_f V g h μ	-509.1514	Simplified genetical path model
16	μ_m μ_f V h μ	-510.0310	
17	μ_m μ_f V g h β μ	-509.0011	Simplified with independent common environment
18	μ_m μ_f V h β μ	-509.2089	

NOTE

Models marked with a "*" are the equilibrium models.

the parents produced significantly more likely solutions than models without it. The importance of an independent environmental component compared with a common environment depending on the parents was confirmed by inspection of the relative magnitudes of the paths p and e.

The most likely biometrical-genetical model was number five, which specifies specific environmental variation, additive genetical variation and a covariance between all individuals living together. The likelihood for this model does not differ significantly from that of our baseline model, number 17. Since this model summarises the pattern of variation using five parameters and model 17 uses seven parameters, we might consider model 5 to be a more satisfactory and parsimonious explanation of individual differences in Radicalism. However, the way in which the common family covariance has been specified does not allow us to achieve a separation between the effects of the common environment and assortative mating, since both these would contribute to the E_2' term. Therefore, we might prefer one of the path models as a more useful way of looking at the data.

The full genetical path models (models 11-14) showed us that the data are consistent with additive genetical variation. There was no reason to suppose, from comparing the likelihoods of model 11 with 12 and 13 with 14, that *g* was significantly different from 0.5 i.e. this path represents only additive effects. We also saw that any effect of cultural transmission was small and that there was no evidence (from comparing models 11 with 13 and 12 with 14) that the population was not in equilibrium.

Given that the path, w, and the genotype-environment correlations are so small, have we any reason to suppose that they are significantly

greater that zero i.e. is the effect of cultural transmission producing significant variation? We fitted models 15 and 16 with genetical variation and no assortative mating to test this possibility. Models 15 and 16 do not differ significantly, again confirming that $g = \frac{1}{2}$ and that genetical variation is additive. Thus in order to determine whether there is any significant effect of cultural transmission we compared models 14 and 16. It turns out the difference in likelihood between these models is small and that the maximum log likelihoods of the five parameter and more complex eight parameter model do not differ significantly. Thus, there is no evidence that variation produced by cultural transmission is needed in order to explain individual differences in Radicalism.

However, this does not mean that there is no common environmental component and so we fitted models in which we specified genetical variation, assortative mating a shared environment not dependent upon the parents. The likelihoods for these models are given in Table B63 next to models 17 and 18. Again there is no evidence that g differs from 0.5 and we consider that model 18 is the most appropriate of this pair. Using the likelihood ratio criterion we tested whether model 18 is more likely than model 16. It turns out that $\chi^2_1 = 1.6442$, which is not significant. Again, we have no evidence for significant between families environmental variation.

Thus, in the final analysis we can explain variation between individuals in their Radicalism scores in terms of unreliability of the scale of measurement, specific individual experiences and additive genetical variation. The between families component found using the twins can be attributed to the effects of assortative mating in the

population. The importance of assortative mating is emphasized by the magnitude of the marital correlation ($\mu = 0.3893$), which is close to that calculated from the data in Section 5, and by the correlation of 0.2636 between the genes of the parents.

Estimates of the means and total variance are also similar to those found previously. Since there is no genotype-environment covariance, the proportion of the total variation produced by additive genetical effects is given by h^2. Thus additive variation accounts for 67.7% of the total variation and specific environmental variation for only 32.3%. These figures agree closely with those estimated in Section 6 from the twin data by assuming that the between families component could be explained purely in terms of assortative mating. We therefore have no evidence for cultural differences between families which cannot be explained in terms of their genotypes, and no evidence for significant cultural transmission. How the determination of man's cultural background affects variation cannot be studied in these data. We could turn to another trait to study this problem. Alternatively, more data collected on a larger sample of adopted individuals, might prove informative, since the value of $\beta = 0.2$ is large although not significant in this study.

5.5 The Final Model for Toughmindedness

The story for Toughmindedness is not so straightforward as that for Radicalism as we can see from Table B64, which summarises the models we have fitted and their corresponding likelihoods. The most likely model, which will act as our baseline model against which to compare

TABLE B 64: SUMMARY OF MODELS FOR TOUGHMINDEDNESS

MODEL	PARAMETERS	LIKELIHOOD
1	$\mu_m \mu_f$ V r	1292.5851
2	$\mu_m \mu_f$ V r μ	1309.0122
3	$\mu_m \mu_f$ V p	1263.5794
4	$\mu_m \mu_f$ V p μ	1271.9892
5	$\mu_m \mu_f$ V p α	1310.6675
6	$\mu_m \mu_f$ V p α μ	1314.7409
7	$\mu_m \mu_f$ V_A g h w s μ	1329.1420
8	$\mu_m \mu_f$ V_A h w s μ	1327.0973
9	$\mu_m \mu_f$ V g h μ	1324.7638
10	$\mu_m \mu_f$ V h μ	1322.7811
11	$\mu_m \mu_f$ V g h β μ	1325.7574
12	$\mu_m \mu_f$ V h β μ	1324.8996

the adequacy of other models is model 7. In the case of Toughmindedness, the full non-equilibrium genetical path model, with g not equal to 0.5, is the most likely model with L = 132 9.1420. Model 8, where $g = \frac{1}{2}$, is the next most likely model, but using the likelihood ratio criterion we see that $\chi^2_1 = 4.0894$, $2\% < p < 5\%$, indicating that model 8 provides a significantly worse explanation of the data than model 7. Since the likelihoods of all the other models are lower than that of model 8, these too fail to explain the observed variation as well as model 7. A purely environmental path model cannot account for individual differences in Toughmindedness when we consider all the data jointly. Yet in the twin data, a simple environmental model adequately explained the observed variation. How can we account for this? In Section 8.4.7 we discussed how model 7 might be more likely than other models because it does not constrain equality of the total variances in parents, natural children and adopted children. This seems possible since the total variances are heterogeneous ($\chi^2_2=19.97$).

Another possibility is that variation is really all environmental, but we have specified the wrong model for the environment. Further work is needed to devise more sophisticated environmental models and test these. However, for the present we must conclude that this genetical model is the most appropriate for our data despite the fact that we detected no genetical variation in the twins. This model breaks the constraint that $g = \frac{1}{2}$, if genetical variation is additive. This may indicate non-additive effects and complex interactions in the data which were not detected in the twins. This seems feasible since detection of non-addivity using twin data is notoriously difficult (Eaves, 1972). It has also been shown in several cases (Eaves, 1977) that significant

non-addivity may be detected using family data which was not apparent for twin data. Thus on our present evidence the most likely model for Toughmindedness seems to be one in which there is genetical variation, additive and non-additive variation, assortative mating, and a substantial environmental component dependent on the parents, which suggests the possibility of cultural transmission. However, these are only tentative conclusions and cannot be taken seriously until further evidence has been collected. Environmental non-additivity, sex interactions or differences in environmental components between groups might provide as likely an explanation if we fitted a model including their effects.

SECTION 9: DISCUSSION

Our overall conclusion from this study of social attitudes confirms earlier results in that a simple picture emerges for Radicalism which is consistent with previous work, where Toughmindedness presents something of an enigma. There was no evidence for genetical variation or complex sex effects on the variances in the twin data for Toughmindedness and this led us to suggest that the trait assessed as Toughmindedness by the Wilson-Patterson Attitudes Questionnaire was not the same as that analysed by previous workers (Eaves and Eysenck, 1974; Hewitt 1974; Martin 1976; Eaves, 1977), whose results were fairly similar. Some evidence was found for this by examining the structure of the questionnaire used in the present study. Our trait of "Toughmindedness" clearly has a different mechanism producing differences between individuals than that of the earlier workers. A picture consistent over twin and adoption studies for the genetical and environmental determinants of this trait has not emerged. The twin data shows the importance of a between families environmental component but provides no evidence for significant genetical variation. The importance of the common environmental component is confirmed for all the data jointly and can be resolved in terms of a model of cultural transmission, since there is evidence for genetical variation when we consider the combined data. Clearly a lot more work needs to be done on Toughmindedness before we can be at all sure of the causes of variation. The trait needs to be defined more precisely, so that we can ascertain the differences between our trait and "Toughmindedness" as measured by previous workers. More family and adoption data is clearly needed in order to provide information on relationships about which we

know little at the moment. Then we can attempt to specify more realistic environmental models for comparison with those we have already tried. We can also explore further the suggestion we had from this study of non-additivity or complex interactions. We know, for example, that there were sex interactions with age which do not appear to be influencing our model fitting but which we would like to study in more detail. We could fit the model of Eaves (1977) which allows for the correlation between relatives to decline with age, following a simple decay curve. This might provide some leverage on the problem of sex interactions in Toughmindedness. It would also be interesting to fit dominance to family data, although dominance could not be detected in the twins. Its effects, or sex interactions might explain the non-additivity we have found. The analysis of Toughmindedness has thus left us with many questions and our main conclusion is that we have an interesting trait on which more work needs to be done.

Radicalism has provided far more clear cut results, which agree closely with those of previous workers. Estimates of specific environmental variation, additive genetical variation and a between families component, $\hat{B}$, are very similar to those obtained in older studies (see Eaves and Eysenck, 1974; Hewitt, 1974; Martin, 1976 and Eaves, 1977). Analagous models to the E_1D_RB model in twins were fitted to the adoption data and the combined data and it was found that twin and adoption studies were homogeneous as far as the E_1, D_R and different E_2 parameters were concerned. Results for Radicalism are thus consistent over different questionnaires at different times and using different family groupings.

Fitting genetical path models to the data showed that the between families component, B, could be interpreted purely as a result of

assortative mating. There was no convincing evidence for environmental differences between families dependent upon the phenotype of the parents, and therefore, no cultural transmission, as defined by Eaves (1976b) or Cavalli-Sforza and Feldman (1973), was found. There was a suggestion that a common environment independent of the parents might be relevant, but its effect was not significant.

Once again, further work needs the collection of more data. In order to provide a more powerful test for a common environment independent of parental phenotype, more adopting parent-child, adopted-adopted child, adopted-natural child and natural sibling pairs are needed. However, other work on this data set would still be desirable despite its deficiencies. Having shown that there is no cultural transmission and that the between families component, $\hat{B}$, can be interpreted in terms of assortative mating, it would be interesting to fit Fisher's model of assortative mating to the data. It would also be useful to fit a genetical path model with no assortative mating to the data. It would also be useful to fit a genetical path model with no assortative mating but with an environmental covariance, although the results from Section 7, suggest that this would be less likely than the genetical model with assortative mating only. Again it would be relevant to look for non-additive effects such as dominance or age effects in view of the recent work of Eaves (personal communication). Another desirable addition to the present work would be to extend Eaves (1976b) model of cultural transmission to include the effects of assortative mating, which was unfortunately not within the scope of this study.

This study has also proved fruitful in extending the methodology for analysing human variation. The unusual combination of maximum likelihood

estimation and pedigree analysis with path model techniques has proved a useful tool in analysing these data. The problems of unbalanced pedigrees discussed in Section 7 have been successfully overcome and maximum likelihood estimates of the parameters of complex models have been obtained, without losing any information by having to omit some types of pairs of individuals from the analysis because there were so few of them. However, the pedigree analysis approach still needs further work. Its chief disadvantage is in its lack of any test of the "goodness-of-fit" of the model. It is useful to be able to pick out the most likely of a series of models. But ultimately we want to be able to determine whether any given model gives an adequate account of the observed variation. Otherwise we may never know about certain important components of variation in our data. Suggestions for ways in which a test of significance might be devised were given in Section 7.

When satisfactory tests of significance have been formulated for the pedigree analysis approach, it should prove the ideal tool for analysing human variation. The need for balanced family structures will disappear and advances in the genetics of human behaviour will no longer have to rely on "nature's experiments", such as twins, which may turn out to be special cases and mislead us about the causes of variation in the population as a whole. The use of pedigree analysis will allow data to be usefully collected on families as they naturally occur, no matter what their structure. Thus random sampling of the variation in any population will be easier and collection of large bodies of data will be practicable and cheaper since we will no longer be looking for rare individuals. Thus, we will be able to make powerful tests of many hypotheses for a variety of different traits and arrive at unequivocal answers about the causes of human variation.

PART C

A STUDY OF POWER AND BIASES IN HUMAN BEHAVIOUR GENETICS

INTRODUCTION

There are several reasons why the work to be discussed in the final part of this thesis was undertaken. Problems which arose during the analysis of the twins studies, particularly the Georgia Twin Study, suggested that it would be useful to know, before embarking on a long and expensive programme of data collection, the optimal experimental design and the number of individuals needed to detect effects of interest and to discriminate between alternative hypotheses about causes of variation in the population. Where the optimal experimental design cannot be used, because of time, expense or the rarity of the individuals required or when analysing data collected by others, then effects of interest may be confounded, may not be statistically significant or may, if we wrongly accept an inappropriate model, be biased by other components of variation than those included in the wrong model. In such cases, it would be desirable to know precisely the biases which may occur.

These problems have not, until recently received widespread attention, but have been considered by members of my own department in experimental organisms (Kearsey, 1970; Mather, 1974) and in man (Eaves, 1969, 1970, 1972, 1977; Martin, 1976).

This work continues that above by considering the power of the test for detecting certain effects and discriminating between alternative hypotheses and the precise estimation of biases introduced where an inappropriate model is accepted.

In the Georgia Twin Study, effective discrimination between the simple genetical and simple environmental models could not be made for several measures of ability and it was suggested that the sample size

was too small to make rejection of the simple environmental model possible given the heritabilities of these particular tests. The genetical model was preferred for other tests, also purporting to measure ability, because an environmental model was unable to explain variation. This raises the problem of how big a sample must be in order to reject a false model at a particular level of probability with a given a given degree of confidence. The power of the test for rejecting false models is discussed by Martin (1976), Eaves (1977) and Eaves, Last, Martin and Jinks (1977), in relation to the classical twin study of monozygotic and dizygotic twins reared together. Martin (1976) assumed various proportions of each type of twin, and generated four statistics (the between and within pairs mean square for MZ and DZ twins), for a series of "populations" with different causes of variation. He considered populations with different combinations and amounts of the following parameters: E_1, E_2, D_R, M (additive variance produced by assortative mating) and H_R. Then, by fitting false models to these populations he obtained the non-centrality parameter of the non-central chisquare distribution of the observed deviations, and from this was able to calculate the number of pairs of twins needed to reject these false models, at the 5% level in 95% of cases, using the chisquare criterion.

Two of his calculations were relevant in the context of the Georgia Twin Study. When he generated mean squares assuming that $E_1 = E_2 = 0.5$ (using a total variance of 1.0) and fitted the E_1D_R model to these "data", he found that 430 pairs of twins (i.e. 215 pairs each of MZ and DZ twins) were needed before the E_1D_R model would be rejected at the 5% level in 95% of cases. Conversely, when the E_1E_2

model was fitted to data assuming that $\frac{1}{2}D_R = E_1 = 0.5$, he found that 640 pairs were needed for rejection of the false model. The Georgia Twin Study contained 364 pairs of twins. Therefore, it could have been predicted that discrimination between the two simple models would be difficult if the heritability of any of the tests was approximately 0.5. Indeed, for those tests where both the E_1E_2 and E_1D_R models fitted the data, the narrow heritability obtained by fitting the E_1D_R model was about 0.5. Thus, since one of the main aims of the Georgia Twin Study was to detect and compare genetical variation in Blacks and Whites, more data should have been collected.

Obviously then, before embarking on any major study, it would be useful to know the results of power calculations, such as those performed by Martin, so that data on an adequate number of individuals could be collected to be reasonably certain of discriminating between the hypotheses that the study is designed to test. Collection of data on fewer individuals that those indicated by the power calculations would clearly be a waste of time and resources.

An allied problem to that of discriminating between alternative hypotheses, concerns the power of the test for detecting particular effects. Given that a particular effect produces variation in a population, the question arises as to how many individuals would be needed for that effect to be significant at the 5% level in 95% of cases in any given experimental design. Kearsey (1970) calculated the sample size needed for the detection of dominance with four different crossing schemes that may be used in experimental organisms, in order to determine the most efficient design. Similar work by Eaves (1972) compared the power of the test for detecting dominance in man in three

different data sets which in theory allow the separation of additive and non-additive components of variation. He showed that sample sizes needed for the detection of dominance variation in human data, using second degree statistics are considerably larger than those normally employed.

Problems of power in a study of MZ and DZ twins reared together, in which the zygosity was not known, were considered by Eaves and Jinks (1972). In this study, the intraclass correlations were determined using the opposite-sex pairs (to provide an estimate of DZ twin correlation) and the known proportion of the two types of twin. They showed that about three times as many pairs of twins would be needed to obtain the same power as in a sample where the zygosity was determined directly using genetic markers or a questionnaire. The methods used by them to determine the power of the test for a genetical component, based on the difference between transformed correlations of MZ and DZ twins, were used in the Georgia Twin Study to determine the number of pairs needed to reject the null hypothesis that there is no genetical component of variation at the 5% level in 95% of cases, given different levels of broad heritability and in the presence and absence of assortative mating. This number was greater than the sample size of the Georgia Twin Study (364 pairs) unless the broad heritability was 0.8 or higher, no matter what the mating system. This showed that the failure to detect significant genetical variation for several measures of ability could not be taken as proof of its absence since the power of the test was too low for those tests whose MZ intraclass twin correlations were below 0.8.

In more recent work, Klein (1974) presented Tables of the power for estimates of heritability and "genetic correlation" obtained from four relationships: offspring-mid-parent, offspring-single parent, full sibs and half sibs. He also showed that large samples are required for the comparison of heritabilities between populations, estimated from each of these relationships.

The problems of the power for rejecting false hypotheses and for detecting genetical and environmental components of variation will be discussed in the following pages. However, the main purpose of the analyses, was to consider the biases that will be introduced into parameter estimates when a false model is wrongly accepted as an adequate explanation of variation. This is likely to happen when the power of the test for detecting certain effects present in a population is too low.

Eaves (1970) considered this problem for the case of dominance, since his power calculations showed that the detection of dominance is likely to be difficult, unless sample sizes are large. Given that diminance is present in a population, he asked how dominance will bias the estimates obtained from fitting false models, which may be adequate to account for the observed variation. He showed, for example, that in studies of twins reared together, when the $E_1E_2D_R$ model is fitted to a population in which E_2 and H_R are both present, then

$$\hat{D}_R = D_R + \tfrac{3}{4} H_R$$

and

$$\hat{E}_2 = E_2 - \tfrac{1}{8} H_R$$

In studies of twins reared together, it is not possible to separate the contributions of E_2 and dominance. Thus, fitting an $E_1D_RH_R$ model, when E_2 and H_R are present, then:

$$\hat{D}_R = D_R + E_2$$

and

$$\hat{H}_R = H_R - 8E2.$$

Thus, E_2 and H_R tend to balance one another and if both are present in a population negative estimates of E_2 or H_R may be obtained depending on their relative magnitude. Such studies of biases can obviously prove useful in interpreting the possible causes of biologically meaningless results.

Mather (1974) derived expectations for the components of variation in terms of three types of epistasis and showed how epistatic interactions will affect estimates of D_R and H_R when allele frequencies are not equal.

The work of Eaves and Mather emphasises how two different sources of non-additive variation can bias estimates of genetical and environmental effects. Here, the biases introduced by genotype-environment interaction (GXE) and genotype-environment covariation (CovGE) will be specified. There is a substantial literature claiming the intractability of GXE and CovGE (e.g. Block and Dworkin, 1974a, b; Layzer, 1974; Linn, 1974; Moran, 1973). However, the specification of these effects, their analysis and their biological significance is still the subject of much confusion. The theory and analysis of GXE in species other than man has been the subject of much research (e.g. Haldane, 1946; Mather and Jones, 1958; Bucio-Alanis et al, 1969; Jinks and Perkins, 1970; Jinks and Connolly, 1975; Mather and Caligari, 1975) and attempts have been made to specify empirically CovGE in man (e.g. Cattell, 1960; Loehlin, 1965). Jinks and Fulker (1970) systematically treat the principles underlying the detection and analysis of GXE and CovGE in man. More recent work by Eaves (1976a,b) has provided a satisfactory theoretical formulation of different sources of CovGE.

However, a number of authors suggest that GXE and CovGE preclude any worthwhile analysis of individual differences in human populations (e.g. Moran, 1973; Layzer, 1974; Lewontin, 1974; Feldman and Lewontin, 1975). They express the view that GXE and CovGE are important in determining individual differences in behaviour, but cannot be detected and separated from other components of variation. Thus, attempts to estimate and interpret other population parameters may be seriously in error. It is suggested that the biases introduced generally lead to inflation of heritability estimates. The problems of detecting and estimating genotype-envirohment interaction and genotype environment covariation have been adequately answered in the work mentioned above. Whether GXE and CovGE are widespread and produce a substantial amount of variation in behaviour in human populations can only be determined by practical investigation of a number of traits using the methods already available. In many cases, the presence of GXE and CovGE will lead to failure of simple models. However, the amount of systematic GXE and CovGE which might remain undetected and the biases this would introduce into parameter estimates can be investigated theoretically.

Plomin et al (1977) make some attempt to assess how GXE and CovGE may bias estimates of genetical and environmental components of variation in twin data, but not in adoption data. However, CovGE produces biases in both types of data. The weakness of their study is that they only suggest what biases may occur. They make little attempt to quantify the biases or deal with the problem in a systematic way.

In order to satisfy criticisms that GXE and CovGE seriously bias parameter estimates and lead to overestimation of the heritability, we will attempt a systematic treatment of the problems discussed by Plomin et al. We will determine under what conditions GXE and CovGE will remain

undetected by performing a series of power calculations and also calculate precisely how different parameters will be biased in a variety of situations. Similar analyses will be performed for variation produced by additive effects and dominance in order to confirm and extend earlier work (Eaves, 1969, 1970, 1972; Martin, 1976).

The results of this study, which are presented in a series of reference Tables in Appendix G, are too extensive to be presented in detail. The reader may extract from the Tables the information most pertinent to his particular problem or interest. However, those findings that are relevant to the arguments about whether GXE and CovGE preclude worthwhile analysis of individual differences in human populations will be discussed.

Part of this work is included in a paper which discusses in detail problems of non-additivity and genotype-environment covariance and attempts to provide a comprehensive framework of theory and method in which these and other contributions to individual differences can be critically assessed. This paper (Eaves, Last, Martin and Jinks, 1977) is included as Appendix F.

SECTION 1: THE "DATA" SETS

1 THE EXPERIMENTAL DESIGN

In order to determine the power for detecting genetical and environmental effects, the power for rejecting false hypotheses and the biases produced by fitting the "wrong" model to particular hypothetical populations, we will consider only one experimental design. In practice a variety of designs are used and ideally we would like to repeat the work to be described here for a variety of the most commonly used designs.

The experimental design adopted here is rather more demanding than those usually employed in studies of human behaviour. It has the disadvantage that it requires certain rare types of individual e.g. monozygotic twins reared apart, and is, therefore, unlikely to be used in practice. However, it is extremely useful in our context because it allows us to separate all effects of interest and, to make powerful tests for them.

The mean squares from the ideal study to which we shall fit models are derived from the following individuals:

Monozygotic twins reared together;

Dizygotic twins (full-sibs) reared together;

Monozygotic twins reared apart;

Dizygotic twins (full-sibs) reared apart;

Unrelated individuals reared together; and

Singletons, reared by their natural parents.

Eleven mean squares can be derived from this study i.e. a between and within pairs mean square for each of the twin types and the total variance of singletons. We could, therefore, fit models including up to ten

parameters and still have one degree of freedom left for testing the adequacy of the model. Thus the complexity of the model we can fit will not be limiting and, in fact, the most complicated models fitted were those involving five parameters. In the classical twin study considered by Martin (1976), only three parameters can be fitted and components of variation such as dominance and the common environmental effect cannot be separated.

In order to specify certain models, it is necessary to have information about the rearing conditions of those related individuals who have been reared apart. For the purposes of this investigation, we will assume that such individuals have been reared as singletons in randomly chosen foster homes. Eaves et al (1977) discuss the effect of placement on family resemblance.

2 GENERATING THE "DATA"

2.1 The Populations

The analyses will be performed for a series of populations with different sources of environmental and genetical variation. There are five main areas of interest: additive effects, dominance, genotype-environment interaction, genotype-environment covariation produced by cultural transmission (Eaves, 1976b) and sibling effects (Eaves, 1976a). Each of these effects will be considered in combination with other sources of variation. The total of fourteen populations is given in Key B in Appendix G. Within each of the five main areas, the causes of variation are of increasing complexity.

2.2 General Considerations for Generating the Populations

Each population type may vary considerably in the proportion of the total variation accounted for by each of its parameters. We will consider a series of possibilities for each population, in order to demonstrate how power changes depending upon the relative magnitudes of different components of variation.

The expected values of the second degree statistics obtained from the study described above will be produced for a variety of systems of causation, which may be represented by various combinations of nine parameters. The contributions of these parameters to the mean squares are given in Table C1.

TABLE C1: CONTRIBUTION OF POPULATION PARAMETERS TO ELEVEN MEAN SQUARES

MEAN SQUARES	EXPECTATIONS								
	E_1	E_2	D_R	M	H_R	I	C	D_R''	D_R'
Between MZT	1	2	1	1	$\frac{1}{2}$	1	2	1	2
Within MZT	1	.	.	.	.	.	.	.	.
Between DZT	1	2	$\frac{3}{4}$	1	$5/16$	$\frac{3}{4}$	2	$\frac{3}{4}$	$1\frac{1}{2}$
Within DZT	1	.	$\frac{1}{4}$	.	$3/16$	$\frac{1}{4}$	.	$\frac{1}{4}$	$-\frac{1}{2}$
Between MZA	1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$	.	.	.
Within MZA	1	1	.	.	.	$\frac{1}{2}$	.	.	.
Between DZA	1	1	$\frac{3}{4}$	1	$5/16$	$\frac{1}{2}$	.	.	.
Within DZA	1	1	$\frac{1}{4}$	.	$3/16$	$\frac{1}{2}$	.	.	.
Between UT	1	2	$\frac{1}{2}$	1	$\frac{1}{4}$	$\frac{1}{2}$	.	$\frac{1}{2}$	1
Within UT	1	.	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	.	$\frac{1}{2}$	-1
Singletons	1	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1	.	.

In order to produce numerical expectations for the eleven mean squares in any population, we must specify values for its parameters. These are arbitrary, but some sets of values are more consistent with the the existing body of information about the causes of variation in human populations. In all cases, we will start with a value of 225 for the total variance (σ^2_T). Certain effects will lead to a change in σ^2_T when we fix the population values (e.g. CovGE produced by sibling effects). We will consider two general levels of broad heritability of 0.5 and 0.9, which roughly correspond to values found for personality traits and social attitudes on the one hand and measures of ability on the other. Thus the total environmental variation (E) will be either

$$E = \tfrac{1}{2} \times 225 = 112.5$$

or

$$E = \frac{1}{10} \times 225 = 22.5$$

Even in cases where genotype-environment covariance leads to a change in σ^2_T, these values for E will be used. The contributions of E_1 and E_2 to E are determined empirically for any real set of data. However, we will assume that $E_1 = E_2$ throughout so that the only possible values for E_1 and E_2 are:

$$E_1 = 112.5 \text{ or } E_1 = E_2 = 56.25$$

and

$$E_1 = 22.5 \text{ or } E_1 = E_2 = 11.25$$

The particular values taken by the other population parameters, vary depending on the particular combination of causes of variation, and will be described for each population in turn. The values of the parameters are summarised in the Tables in Appendix G.

2.3 Population 1: E_1D_R

In the simplest, random mating population, there is specific environmental variation and additive genetical variation. Two cases only will be considered, according to the general criteria given above. These are:

a $\frac{1}{2}D_R = E_1 = 112.5$

b $\frac{1}{2}D_R = 9E_1 = 202.5$ i.e. $E_1 = 22.5$

2.4 Population 2: $E_1E_2D_R$

In this population, there is additive gene action and environmental variation within and between families. When $h^2_B = 0.5$, we have $\frac{1}{2}D_R = 112.5$ and $E_1 = E_2 = 56.25$. In the second case, $h^2_B = 0.9$ and we have $\frac{1}{2}D_R = 202.5$ and $E_1 = E_2 = 11.25$.

2.5 Population 3: E_1D_RM

There is specific environmental variation and additive genetical variation as in Population 1, but now there is also additive variation produced by assortative mating. As in population 1, we will consider two cases where

a. $\frac{1}{2}D_R = E_1 = 112.5$

and b. $\frac{1}{2}D_R = 9E_1 = 202.5$

Assortative mating leads to an increase in additive genetical variance until at equilibrium, the contribution of the additional variation is

given by $\frac{1}{2}(A/(1-A))D_R$, where A is the correlation between the additive genetical deviations of spouses. By analogy with D_R, we will define M as $(A/(1-A))D_R$, so that the contribution of assortative mating to σ^2_T is $\frac{1}{2}M$.

Assuming Fisher's model of assortative mating (Fisher, 1918), $A = \mu h^2_N$, where μ is the correlation between the phenotypes of spouses. In order, to generate the populations, we will assume values for A which are likely to represent values found for measures of ability (Vandenberg, 1972) and lie within the range of values found for other behavioural traits.

The values of A chosen were 0.25 and 0.45 for the two cases. From these values of A, M was found as $(A/(1-A))D_R$ which was 75.0 when $\frac{1}{2}D_R / \sigma^2_T = E_1$; and 331.4 when $\frac{1}{2}D_R / \sigma^2_T = 9E_1$. This procedure enables reasonable values of M to be calculated for given D_R, E_1 and A, remembering that $A = h^2_N \mu$ at equilibrium. We may calculate equilibrium values of h^2_N, μ and σ^2_T. We can find $\sigma^2_T = E_1 + E_2 + \frac{1}{2}D_R + \frac{1}{2}M = 262.5$ for case 1 and $\sigma^2_T = 390.7$ for case 2.

Thus assortative mating can lead to a substantial increase in variation.

We calculate $h^2_N = \frac{1}{2}D_R + \frac{1}{2}M / \sigma^2_T$, which gives us:

a. $h^2_N = 0.57$

b. $h^2_N = 0.94$

The marital correlation can now be estimated as $\mu = A/h^2_N$ to give:

a. $\mu = 0.44$

b. $\mu = 0.48$

The final parameters for these populations are summarised in Table C2.

TABLE C2: PARAMETERS OF SIMPLE POPULATIONS WITH ASSORTATIVE MATING

PARAMETER	POPULATION			
	3a	3b	4a	4b
E_1	112.5	22.5	56.25	11.25
E_2	-	-	56.25	11.25
D_R	225.0	405.0	225.00	405.00
M	75.0	331.4	75.00	331.36
σ^2_T	262.5	390.7	262.50	390.68
h^2_N	0.57	0.94	0.57	0.94
A	0.25	0.45	0.25	0.45
μ	0.44	0.48	0.44	0.48

2.6 Population 4: $E_1\ E_2\ D_R M$

This is similar to the previous population, but we now have two sources of environmental variation. Since M is defined purely in terms of D_R and A, the only parameters which differ between this population and the previous population are E_1 and E_2. The two cases where $h^2_N = 0.57$ and $h^2_N = 0.94$ are summarised in Table C2.

2.7 Population 5: $E_1 D_R H_R$

This population contains specific environmental variation and additive genetical variance, as in all previous populations, but now contains dominance. We will consider two levels of heritability and two levels of dominance to provide four combinations. As before when $h^2_B = 0.5$, $E_1 = 112.5$ and when $h^2_B = 0.9$, $E_1 = 22.5$.

We will consider the cases where dominance is complete i.e. where the ratio, $\sqrt{D_R/H_R}$ (i.e. dominance ratio when allele frequencies are equal) is 1 and where dominance is intermediate. For intermediate dominance, a dominance ratio of 0.5 is used.

In order to perform the calculations, we let $\sqrt{R} = D_R/H_R$.

Therefore, H_R is given by:

$$H_R = D_R . R^2$$

The total genetical variance, $\sigma^2_G = \frac{1}{2}D_R + \frac{1}{4} H_R$. This may be partitioned into components due to additive genetical and dominance variation:

$$D_R = \frac{\sigma^2_G}{\frac{1}{2} + \frac{1}{4} R^2}$$

$$H_R = D_R . R^2$$

For $h^2_B = 0.5$, given equal allele frequencies, a dominance ratio of 1 will give $D_R = H_R = 150$ and a dominance ratio 0.5 gives $D_R = 200$ and $H_R = 50$. In each case the contribution of D_R and H_R to the total variation, given by $\frac{1}{2}D_R + \frac{1}{4}H_R$, is 112.5. Similarly, when $h^2_B = 0.9$, we find $D_R = H_R = 270$ for a dominance ratio of 1 and $D_R = 360$ with $H_R = 90$ for a dominance ratio of 0.5.

2.8 Population 6: $E_1\ E_2\ D_R\ H_R$

This is similar to the previous population, but there is now between families environmental variation. The calculations involved in obtaining D_R and H_R for the two levels of heritability and dominance ratio are identical to those in 2.7, yielding the same values as before. The only difference is that now when $h^2_B = 0.5$, $E_1 = E_2 = 56.25$ and when $h^2_B = 0.9$, $E_1 = E_2 = 11.25$.

2.9 Population 7: $E_1 D_R H_R M$

In this population both dominance and assortative mating are present. The values for D_R and H_R are calculated as in 2.7, and are, identical to those given for the $E_1 D_R H_R$ population for the four cases. Variation produced by assortative mating is added to the total variation as in 2.5. Calculating M as before, we find four different values for M:

a. $M = 30.00$ when $E_1 = 112.5$ and $R = 1$

b. $M = 115.71$ when $E_1 = 22.5$ and $R = 1$

c. $M = 57.14$ when $E_1 = 112.5$ and $R = \frac{1}{2}$

d. $M = 240.00$ when $E_1 = 22.5$ and $R = \frac{1}{2}$

(using $A = 0.156, 0.239, 0.197, 0.261$, respectively).

This gives new values of σ^2_T of 240.00, 282.86, 253.57 and 345.

The true narrow and broad heritabilities are, therefore:

	h^2_N	h^2_B
a.	0.375	0.531
b.	0.682	0.920
c.	0.507	0.556
d.	0.870	0.935

The values of the marital correlation in the four populations respectively are: 0.416, 0.350, 0.389, and 0.300).

2.10 Population 8: $E_1\ E_2\ D_R\ H_R\ M$

This population has identical values for D_R, H_R and M as the previous population. However, at the lower and higher levels of heritability, $E_1 = E_2 = 11.25$ and $E_1 = E_2 = 56.25$ respectively.

2.11 Population 9: $E_1\ E_2\ D_R\ I$

This population and the two that follow contain genotype-environment interaction, which is specified using the parameter "I".

This population is randomly mating and there is additive gene action and environmental variation within and between families. Environmental differences between families are independent of genetical differences, but the two sources of variation interact, producing variation due to the interactions of D_R and E_2. The contribution of GXE to the total variance is $\frac{1}{2}$ I. Interactions involving E_1 are not included since these are inevitably confounded with estimates of E_1.

We will consider the two cases where $E_1 = E_2 = 56.25$ and $E_1 = E_2 = 11.25$. The remaining 50 per cent of the variation in the former case and 90 per cent in the latter case will be assigned to the additive effects of genes and their interaction with environmental differences between families i.e. $\frac{1}{2}D_R + \frac{1}{2}I = 112.5$ and 202.5 respectively. Two levels of interaction will be used. In experimental organisms, it is seldom found that more than 20 per cent of the variation can be attributed to GXE. Therefore, values of 20 per cent and 10 per cent of the total variation will be used for GXE in order to examine its effect in extreme and less extreme cases. Ten per cent of the total variation is 22.5 and twenty per cent is 45. Therefore, values of 45 and 90 were given to I. D_R, which is given by $2 \times (225 - (E_1 + E_2) - \frac{1}{2}I)$, takes the values 180 and 360 when $\frac{1}{2}$I accounts for 10 per cent of the variation and $h^2_B = 0.5$ or 0.9 where $h^2_B = \dfrac{\frac{1}{2}D_R + \frac{1}{2}I}{\sigma^2_T}$

When $\frac{1}{2}$I accounts for 20 per cent of the variation, D_R takes the values 135 and 315 respectively.

2.12 Population 10: E_1E_2 D_R I M

Both GXE and assortative mating are found in this population. The

four sets of parameter values involving E_1, E_2, D_R and I which were calculated in the last section are again used. Then variation produced by assortative mating is calculated and added to the total variance as described in 2.5.

Values of A of 0.2, 1.15, 0.4 and 0.35 are used. We can, therefore, calculate M as $(A/(1 - A))\ D_R$, which gives values of 45.00, 23.82, 40.00 and 169.62. Therefore, the total variances become: 247.5, 236.91, 345 and 309.81. The values for h^2_N in this population are calculated as:

$$h^2_N = \tfrac{1}{2}D_R + \tfrac{1}{2}M/\sigma^2_T$$

This gives values of: 0.45, 0.34, 0.87 and 0.78. If we define the broad heritability as:

$$h^2_B = \frac{\tfrac{1}{2}D_R + \tfrac{1}{2}I + \tfrac{1}{2}M}{\sigma^2_T}$$

we obtain values: 0.55, 0.53, 0.93 and 0.93. The marital correlations given as $\mu = A/h^2_N$ become: 0.89, 0.88, 0.92 and 0.90.

2.13 Population 11: E_1 E_2 D_R H_R I

In this population we may vary the broad heritability, the dominance ratio and the amount of GXE interaction. Using $h^2_B = 0.5$ or $h^2_B = 0.9$, dominance ratios of 1 and $\frac{1}{2}$ and GXE accounting for 10 or 20 per cent of the total variation yields eight possible situations. These are given in Table 9B of Appendix G.

In order to generate the population values the following steps were taken:

1. Values of E_1 and E_2 were fixed at either $E_1 = E_2 = 56.25$ or $E_1 = E_2 = 11.25$. This leaves 112.5 or 202.5 to be assigned to the remaining sources of variation.
2. Either 10 per cent or 20 per cent of the remaining variation was assigned to $\frac{1}{2}I$, to produce values of I of 45 and 90 respectively. Subtracting these from the remaining variation leaves 90 for $E_1 = E_2 = 56.25$ when $\frac{1}{2}I = 22.5$ and 67.5 when $\frac{1}{2}I = 45$; and 180 for $E_1 = E_2 = 11.25$ when $\frac{1}{2}I = 22.5$ and 157.5 when $\frac{1}{2}I = 45$.
3. The remaining variation is partitioned between D_R and H_R for the two dominance ratios as described in 2.7. This produces eight pairs of values for D_R and H_R: 120, 120; 160, 40; 90, 90; 120, 30; 240, 240; 320, 80; 210, 210; 280, 70; where the dominance ratio changes fastest, then the amount of GXE and then finally the broad heritability. This is made clear in Table 9B of Appendix G.

2.14 Population 12: E_1 E_2 D_R C

This population contains environmental variation within and between families, additive genetical variation and CovGE arising as a result of the cultural impact of parents on their offspring. Earlier work (Cattell, 1960; Loehlin, 1965) made no attempt to specify a theoretical basis for the causes of CovGE, merely demonstrating the contributions of the empirical parameters, E_2 and CovGE, to the total variation. Eaves (1976a, b) has described several mechanisms by which CovGE may arise. One mechanism is cultural transmission (Eaves, 1976b) whereby phenotypic variation in the parents is perpetuated culturally as well as genetically in the offspring generation. The parameters E_2 and CovGE can thus be reformulated:

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

and

$$2CovGE = \frac{2b}{1-b} \left(\tfrac{1}{2} D_R\right) \text{ for } |b| < \tfrac{1}{2}$$

where b is the regression of offspring family environment on parental phenotype. The symbol "C" is used to denote 2CovGE throughout. We consider only the simplest population where E_1, E_2, D_R and C are present. Thus, common environmental effects not dependent upon the parents are excluded in these analyses, but could occur in principle.

Two possibilities are considered for the values of D_R and E_1. We let $\frac{1}{2}D_R = 2E_1 = 112.5$ or else $\frac{1}{2}D_R = 2E_1 = 202.5$, corresponding to narrow heritabilities of 0.67 and 0.95, in the absence of cultural transmission. If we now assume that cultural transmission is entirely responsible for common environmental effects, values for E_2 and CovGE can be calculated if the magnitude of b is fixed. We chose to consider values of b which were low (b = 0.1) and intermediate (b = 0.25). For b = 0.1, we get $E_2 = 61.61$ and C = 150.00 when $\frac{1}{2}D_R = 2E_1 = 112.50$ and $E_2 = 98.04$ and C = 270.00 when $\frac{1}{2}D_R = 2E_1 = 202.50$. Similarly we find $E_2 = 6.76$ and C = 50.00 when $\frac{1}{2}D_R = 2E_1 = 112.50$ and $E_2 = 10.33$ and C = 90.00 when $\frac{1}{2}D_R = 251 = 202.50$.

2.15 Population 13: $E_1\ D_R\ D_R''\ D_R'$

Populations 13 and 14 both consider another source of genotype-environment covariance, namely sibling effects. (Eaves, 1976a). The genotype-environment covariance in this instance arises because the phenotypes of siblings form a part of the environment of their co-siblings

In population 13, we consider a simple system in which all gene effects are additive and all the between families environmental component is due to the genetical covariance between members of a sibling pair. The total variance of singletons is thus: $\sigma^2_S = \frac{1}{2}D_R + E_1$, since there is no variation due to sibling effects in individuals reared alone. The total variance of singletons is 225 and we will consider two levels of heritability. When $h^2_N = 0.5$, $E_1 = 112.5$ and $D_R = 225.0$. For $h^2_N = 0.9$, $E_1 = 22.5$ and $D_R = 405.0$. Variation produced by sibling effects must now be specified. Eaves (1976a) uses two parameters to summarise this variation. These are:

D_R'' - the variation produced by the contribution of the additive genetical deviations of siblings to the environment of their co-siblings (the "genetic environmental" variance due to sibling effects) and

D_R' - the covariance between genotype and environment produced by the contribution of siblings phenotypes to the environments of their co-siblings.

Table C1 shows the contribution of D_R'' and D_R' to the mean squares for the hypothetical experiment described earlier, which depend upon the degree of relationship between members of the pair contributing to a particular statistic. We may contrast this with the situation when E_2 and CovGE are specified empirically.

For each level of heritability, we will consider both $D_R'' = D_R$ and $D_R'' = \frac{1}{2} D_R$. The specification of D_R' is more difficult. For the purpose of these simulations, we will define

$$\rho_{ge} = D_R' / \sqrt{(D_R D_R'')}$$

Since D_R' is in effect a measure of association between the direct effects of genes on an individuals phenotype and their indirect effect, mediated through the environment, on the phenotype of a sibling. ρ_{ge} is the proportion of loci having both direct effects on the phenotype and indirect effects on the phenotype of siblings, under certain conditions stated by Eaves et al (1977). Values of $\rho_{ge} = 1.0$ and $\rho_{ge} = 0.5$ were used in this study. We now have two variables for each heritability level. When $h^2_N = 0.5$. $D_R = 225.0$ and D_R'' takes either the value 225.0 or 112.5. For each of these possibilities D_R' may take two values depending on ρ_{ge}. Thus we have:

a. $D_R = D_R'' = D_R' = 225$ } $\rho = 1$ } $h^2_N = 0.5$

b. $D_R = 2D_R'' = 225$; $D_R' = 159.1$ } $\rho = 1$ } $h^2_N = 0.5$

c. $D_R = D_R'' = 2D_R' = 225$ } $\rho = \frac{1}{2}$ } $h^2_N = 0.5$

d. $D_R = 2D_R'' = 225$; $D_R' = 79.55$ } $\rho = \frac{1}{2}$ } $h^2_N = 0.5$

e. $D_R = D_R'' = D_R' = 405$ } $\rho = 1$ } $h^2_N = 0.9$

f. $D_R = 2D_R'' = 202.5$; $D_R' = 286.38$ } $\rho = 1$ } $h^2_N = 0.9$

g. $D_R = D_R'' = 2D_R' = 405$ } $\rho = \frac{1}{2}$ } $h^2_N = 0.9$

h. $D_R = 2D_R'' = 405$; $D_R' = 143.19$ } $\rho = \frac{1}{2}$ } $h^2_N = 0.9$

These values assume that the CovGE is positive i.e. that there is co-operation and the high performance of one sib produces an environmental effect enhancing performance of a co-sib. The converse situation of competition is also possible in which the CovGE is negative and the high performance of one sib is achieved at the cost of another. In this case, the same absolute values for $E_1 D_R, D_R''$ are D_R' are used, but D_R' is now negative. This gives a total of sixteen possible situations for a population in which $E_1 D_R$, D_R'' and D_R' are present (see Table 11B).

2.16 Population 14: $E_1\ E_2\ D_R\ D_R''\ D_R'$

This population is similar to the previous one, except that we have now introduced environmental variation between families. The values of D_R, D_R'' and D_R' are identical to those used previously, but now $E = E_1$ is replaced by $\frac{1}{2}E = E_1 = E_2$. Thus $E_1 = E_2 = 56.25$ when $h^2_N = 0.5$ and $E_1 = E_2 = 11.25$ when $h^2_N = 0.9$. The parameters of these populations are given in Table B12 in Appendix G.

2.17 Generating the Mean Squares from the Populations

In order to proceed, we must now generate the eleven mean squares which could be derived in our hypothetical experiment for each situation for each population type. This involves a considerable number of sets of mean squares, but the procedure followed is the same in each case. We will illustrate this by reference to the simplest population $(E_1 D_R)$, when $h^2_N = 0.5$.

The vector of parameters used to generate the statistics is in this case:

$$\begin{pmatrix} E_1 \\ D_R \end{pmatrix} = \begin{pmatrix} 112.5 \\ 112.5 \end{pmatrix}$$

The expected values of the statistics which can be derived from this population in our experiment are obtained by post multiplying the model matrix which can be derived from the appropriate columns of Table C1 by this vector of population parameters:

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 1 & \frac{3}{4} \\ 1 & \frac{1}{4} \\ 1 & 1 \\ 1 & 0 \\ 1 & \frac{3}{4} \\ 1 & \frac{1}{4} \\ 1 & \frac{1}{2} \\ 1 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix} \quad x \quad \begin{pmatrix} 112.5 \\ 112.5 \end{pmatrix}$$

For example, this produces the first expected mean square for MZT as:

$$\begin{aligned} MS_{bMZT} &= E_1 + D_R \\ &= 112.5 + 112.5 \\ &= 225 \end{aligned}$$

and the within pairs mean square for MZT as:

$$\begin{aligned} MS_{wMZT} &= E_1 + (0 \times D_R) \\ &= 112.5 \end{aligned}$$

The mean squares for the wide variety of situations considered in this study are not given, but can easily be derived for any situation of interest by the procedure used above, using the appropriate vector of population parameters which can be found from Tables 1B to 12B in Appendix G and the model matrix derived from the appropriate columns of Table C1.

SECTION 2: THE ANALYSES

1. FITTING MODELS

The procedures to be described in this Section are general and will be applied to each of the "data sets" generated in turn. Population 1 ($E_1 D_R$) will be used to illustrate the analyses. The case where $E_1 = \frac{1}{2}D_R = 112.5$ will be considered.

The three types of calculation to be performed have a common starting point. A model is fitted to the mean squares from a simulated population 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. The principles and methods of weighted least squares are described in some detail in Section 3 of Part A and Appendix E.

Estimates of the parameters of the model are given by $(\underset{\sim}{A}'\underset{\sim}{W}\underset{\sim}{A})^{-1}\underset{\sim}{A}'\underset{\sim}{W}\underset{\sim}{x}$.

where $\underset{\sim}{x}$ is the vector of mean squares

$\underset{\sim}{A}$ is a model matrix

$\underset{\sim}{W}$ is a diagonal matrix of weights

The weights are given by:

$$w_i = \frac{n_i}{2V_i^2} \qquad (i = 1, 11)$$

where V_i is the value of the ith mean square for the first iteration, but is evaluated from the parameter estimates in subsequent iterations. n_i is the degrees of freedom for the ith mean square.

We have described the structure of a hypothetical experiment which produces eleven mean squares but have not yet fixed its size. For the

purposes of these calculations, there are 200 pairs of each twin type and one hundred singletons. The n_i used in calculating the weights were each taken to be 100. This is an approximation since the degrees of freedom between pairs should be one less than those within pairs. This simplification facilitates repetition of the calculations, and has only a trivial effect on the results, since the degrees of freedom are large.

Standard errors of the estimates (σ_θ) be obtained from the covariance matrix of the estimates, which is given by:

$$\underset{\sim}{Z} = (\underset{\sim}{A}'\underset{\sim}{W}\underset{\sim}{A})^{-1}$$

Dividing the estimates ($\underset{\sim}{\theta}$) by their standard errors ($\underset{\sim}{\sigma_\theta}$) yields a vector of normal deviates ($\underset{\sim}{C}$), which are the expected values of $\theta_i / \sigma_{\theta_i}$ for samples of this size and are used to test the significance of the estimates in real data. In our case, the expected values of the normal deviate will be used to assess the power of the test for detecting particular effects.

The significance of the residuals after a model has been fitted to real data is given by

$$\chi^2_{N-k} = (O_i - E_i)^2 W_i$$

where N is the number of statistics

k is the number of parameters

O_i are the observed mean squares

E_i are the mean squares predicted from the parameter estimates

W_i is the weights matrix evaluated from the estimates which minimise the value of χ^2_{N-k}

Normally when fitting models to real data $\mathcal{E}\chi^2_{N-k} = N-k$ when the model fits because the mean squares are estimated with error from the population. However, when fitting a "fa e" model, the mean squares from the hypothetical populations are known exactly and $\mathcal{E}\chi^2_{N-k} = N-k+\lambda$, where λ is the non-centrality parameter of a non-central chisquare with N-k degrees of freedom. It turns out that the weighted residual sum of squares which we calculate by fitting "wrong" models to expected mean squares is approximately equal to the non-centrality parameter and can be used to calculate the power of the test of goodness of fit of false hypotheses.

The final calculations we make will estimate the biases introduced into parameter estimates, when the model fitted is not the true model of variation.

2 THE POWER OF THE TEST FOR DETECTING EFFECTS

Two models were fitted to the meansquares derived from the population to be used as an example throughout. These are the E_1E_2 and E_1D_R models. A number of other models were also fitted to the mean squares from the remaining populations. The full set of models is given in Key C of Appendix G. The most complex model fitted to any set of mean squares was one with the same number of parameters as the causes of variation in the population from which the mean squares were derived. In most cases, more complex models would contain the true causes of variation as a subset of their parameters and it would not make sense to fit them since no residual variation will remain after all the true causes of variation have been fitted because the mean squares from our populations are known without error. In real data fitting a model which is too complex should produce non-significant residuals.

The E_1D_R model is the "true" model for the causes of variation in our example population and a "true model" will be fitted to each population in turn. A series of false models will also be fitted as exemplified by the E_1E_2 model for our example. For more complex populations a whole series of false models are possible and were fitted to the mean squares. These models can be found for each population type in the tables in Appendix G.

When the E_1D_R model was fitted to the mean squares, the variance-covariance matrix was obtained:

$$\underset{\sim}{Z} = (\underset{\sim}{A}'\underset{\sim}{W}\underset{\sim}{A})^{-1}$$

where $\underset{\sim}{W}$ was an 11 x 11 diagonal matrix of weights. The off-diagnonal terms are zero because the mean squares are independent.

The covariance matrix is:

$$\begin{array}{c} \\ E_1 \\ D_R \end{array} \begin{array}{c} \begin{array}{cc} E_1 & D_R \end{array} \\ \begin{pmatrix} 103.6688 & -170.2078 \\ -170.2078 & 671.7273 \end{pmatrix} \end{array}$$

Thus, the estimates and their standard errors are:

$$\begin{array}{c} \\ E_1 \\ D_R \end{array} \begin{array}{c} \begin{array}{cc} & \end{array} \\ \begin{pmatrix} 112.50 & 10.18 \\ 225.00 & 25.92 \end{pmatrix} \end{array}$$

Dividing each estimate by its standard error we obtain a vector of expected normal deviates:

$$c = \begin{pmatrix} 11.05 \\ 8.68 \end{pmatrix}$$

These are the values of c expected for samples of 1100 in an experiment of this design, where individual differences are determined equally by by specific environmental variation and additive genetical variation. Normally a two-tailed test of significance is performed, although the components of variation are expected to be positive if the model is appropriate, because, in practise, we may fit inappropriate models and significant negative estimates would indicate that the model was wrong. Thus, we may determine whether a model provides an adequate explanation of the data by looking at the significance of the residuals, or by deciding whether parameter estimates are consistent with biological theory. As can be seen from the Tables in Appendix G, negative estimates are often

obtained when inappropriate models are fitted and in these circumstances it is impossible to mislead ourselves about the true causes of variation.

The expected values of c show that we would have no difficulty in detecting both E_1 and D_R as effects significantly different from zero in a sample of this size and structure. However, this sample is larger than those frequently used in practise and we may ask how many individuals would be required for these parameters to be significant at the 5 per cent level in, for example, 95 per cent of cases. The method used is outlined in Eaves' (1972) paper.

The expected value of c for any parameter is a function of the square root of the total sample size. Thus, the value of c, c_x, for a sample of size x can be calculated, given the expected value of c_e and the sample size in this experiment:

$$c_x = c_e\sqrt{1100/x}$$

or x may be calculated

$$x = \left(\frac{c_e}{c_x}\right)^2 \times 1100 \qquad \ldots 1$$

We want a value of x, such that the null hypothesis that θ is not greater than zero will be rejected at the 5 level in 95% of cases. Thus, the power of the test is the probability that $\hat{\theta} > 1.96$ given that $\theta > 0$. For this probability to be 0.95, we require sample sizes which would produce an expected value of 1.96 + 1.65 = 3.61 for θ/σ_θ. We may substitute for c in equation 1:

$$x = \frac{13.03}{c_x^2} \times 1100$$

For E_1 we have $c_x = 11.05$ and for D_R $c_x = 8.68$. Therefore, we can calculate x as 117 for E_1 and 190 for D_R.

Thus the sample sizes needed to detect E_1 and D_R, in a population where these are the true causes of variation are by no means prohibitive. Corresponding sample sizes when $h^2_N = 0.9$, are 141 for E_1 and 41 for D_R, so that D_R can be detected with very few individuals. This will not be the case for some of the effects present in particular populations, as we shall see in the discussion, where we confirm, for example the work of Eaves (1972) and show that extremely large numbers of individuals are needed for the detection of dominance in four different population types, where dominance is present in association with a variety of other sources of variation (Tables B5 to B8).

Following the procedure outlined above, we also fitted the false E_1E_2 model to the population in which E_1 and D_R were the true causes of variation (and $h^2_N = 0.5$), producing the following estimates (θ_F), standard errors (σ_{θ_F}) and values for the normal deviate (c_F);

	$\underset{\sim}{\theta}_F$	$\underset{\sim}{\sigma}_{\theta_F}$	$\underset{\sim}{c}_F$
E_1	168.75	12.89	13.09
E_1	56.25	12.67	4.44

The number of individuals needed for these "false" estimates to be significant in a two tailed test at the 5 per cent level in 95% of cases was calculated in the same way as for the "true" estimates. We found that 84 and 728 individuals for E_1 and E_2 respectively would be needed when $h^2_N = 0.5$ and 90 and 205, when $h^2_N = 0.9$. Thus, we can obtain "sensible" and significant estimates for a source of variation

which does not exist in a population, with a reasonably small number of individuals. The number of individuals needed to detect biased estimates in other more complex situations will be discussed later.

3 THE POWER OF THE TEST FOR REJECTING FALSE HYPOTHESES

We have discussed two criteria for determining whether a model is a satisfactory explanation of variation in a population. If any estimate of a variance component is significantly negative, the model is deemed inappropriate immediately, although as we shall see later, when certain estimates are negative, alternative more suitable hypotheses are suggested by this fact. A well known example is that of dominance and common environmental variance. A negative value for estimates of one of these effects in twin studies is strongly indicative that the other is present (Eaves, 1970).

A second criterion often used in determining whether a model is adequate is the significance of its parameters. We are suspicious of models in which certain estimates are not significant even if we have no reason to suppose that the residuals are other than zero. In such cases, we may have a misleading parameterisation of the causes of variation. We shall see that the sample sizes needed to detect sources of variation which are not present in a population are usually so high as to make it unlikely that they will be significant at the 5 per cent level in any experiment of the scale normally adopted. However, as we have seen, in certain cases significant estimates of "non-existent" sources of variation may be obtained. In such cases, our usual criterion of the significance of the residuals in a chisquare test is the only one we have. This brings us to the question of the power of this test for rejecting false models. We have already seen that this is particularly relevant for studies of fairly typical sample sizes where discrimination between simple models of variation (i.e. E_1E_2 and E_1D_R) is not possible (see Part A, the Georgia Twin Study and Martin, 1976). Martin (1976) has already

considered the "Power of the Classical Twin Study", for populations in which the causes of variation are fairly simple. We shall repeat his calculations, for our experimental design and also consider the more complex populations which were specified earlier.

The approach of Martin (1976) is based on the non-central chisquare distribution and was suggested by Lewontin's power calculations for the rejection of the null hypothesis of Hardy-Weinberg equilibrium (Lewontin and Cockerham, 1959). He gives the following proof supplied by Dr. Paul Davies which is based on a paper by Mitra (1958).

Given a set of observed statistics, O_i, their expected values calculated on the basis of a "true" model of variation, E_i, and the expected values calculated on the basis of a false model, F_i, we wish for each i, to test the null hypothesis:

$$H_o : O_i = E_i$$

against the alternative hypothesis

$$H_1 : O_i = F_i = E_i + \mu_i / \sqrt{\delta_i}$$

where μ_i is a deviation from the expected value E_i, δ is the degrees of freedom and $\mu_i = \sqrt{\delta_i}\,(F_i - E_i)$. If δ_i is large and H_o is true, then: $O_i \sim N(E_i, 2E_i^2/\delta_i)$ while if H_1 is true, then:

$O_i \sim N(F_i, 2F_i^2/\delta_i) \sim$ approx. $N(F_i, 2E_i^2/\delta_i(1 + o(\delta_i^{-\frac{1}{2}}))$

where $o(\delta_i^{-\frac{1}{2}})$ denotes a term of the order $\delta_i^{-\frac{1}{2}}$.

Thus, $$\frac{\delta_i(O_i-E_i)}{\sqrt{2E_i^2}} \sim N\left\{\frac{\sqrt{\delta_i}\,(F_i-E_i)}{\sqrt{2E_i^2}}\ ,\ 1 + o(\delta_i^{-\frac{1}{2}})\right\}$$

and the asymptotic power function of $\dfrac{\delta_i(O_i-E_i)^2}{2E_i^2}$

is non-central chi-square with non-centrality parameter: $\lambda' = \dfrac{\delta_i(F_i-E_i)^2}{2E_i^2}$

In general, to test H_o: $\underset{\sim}{O} = \underset{\sim}{E}$ against H_1:

$$\underset{\sim}{O} = \underset{\sim}{F} = \underset{\sim}{E} + (\mu_i/\delta_i)$$

we use $$\chi^2_{n-k} = \sum_i \frac{(O_i - E_i)^2}{2E_i^2}$$

with limiting power function being non central χ^2_{n-k} with non centrality parameter $\sum_i \frac{(F_i - E_i)^2}{2E_i^2}$.

The larger the degrees of freedom, δ_i, the larger the deviations, μ_i, may be from E_i before the distribution departs from non central chisquare.

The non centrality parameter, λ', can be obtained when we fit a "false" model to a set of statistics derived from one of our hypothetical populations. The procedure is the same as that used to obtain chi-square, when we fit models to real data. We defined $\chi^2_{n-k} = (O_i - E_i)^2 W_i$ and since $W_i = \delta_i/2E_i^2$, and our mean squares for the hypothetical populations are known without error, then $\chi^2_{n-k} = \lambda'$.

We can use the value of λ' in conjunction with tables of non-central chisquare to determine the sample sizes necessary reject a "false" hypothesis at a certain level of probability with a given degree of confidence. The procedure is best illustrated by reference to the simple example population containing E_1 and D_R.

When the false E_1E_2 model was fitted to this population, we obtained $\chi^2_9 = \lambda'$ (with 9df) = 46.3611. In order to be 95 per cent certain of rejecting the "false" model at the 5 per cent level, we must ensure that $\lambda = 23.589$, which is the value of $\lambda_{(0.05,\ 0.95,\ 9)}$ tabulated by Pearson and Hartley (1972). The number of individuals required is given by $N = \lambda / \lambda'$ for unit sample size. Since our "experiment was based on 1100 individuals, N becomes:

$$N = \frac{\lambda}{\lambda'} \times 1100 = \frac{23.589}{46.3611} \times 1100$$

$$= 562$$

When the calculations were repeated for the population in which $h^2_N = 0.9$, we found N = 147. Thus, although as we saw earlier very few individuals are required for the estimate of E_2 to be significant at the 5 per cent level, the false hypothesis may be rejected with only a modest experiment when $h^2_N = 0.9$. However, even with the optimal experimental design used here, as many as 562 individuals are required to discriminate between simple environmental and simple genetical hypotheses when $h^2_N = 0.5$. This may be a more realistic level of heritability to consider particularly for personality and social attitudes. Thus it seems that if we are to learn anything useful about the causes of individual differences for many behavioural traits we must be prepared to embark on extensive data collection programmes.

4 CALCULATION OF BIASES WHEN FALSE MODELS ARE FITTED

We have considered three ways in which "false" models may be rejected:

1. Negative estimates of variance components
2. Parameters not significant
3. Residuals significant in a chi-square test

We have also described a method for determining the number of individuals needed to detect particular effects when they are present in a population and we fit the correct model. From the Tables in Appendix G, we see that this number is unrealistically high in several cases (e.g. dominance, genotype-environment interaction). This provides some evidence for the claims of several authors, mentioned earlier, that such effects as genotype-environment interaction and genotype-environment covariation may remain undetected even in the most extensive studies of human variation. We may now ask what basis there is for the accompanying claim that such undetected effects seriously bias our estimates.

The calculations needed in order to estimate the biases which occur when a "false" model is fitted are easily performed. When we fitted the E_1E_2 model to the mean squares derived from a population in which $\frac{1}{2}D_R = E_1 = 112.5$, the parameter estimates were obtained as $(A'WA)^{-1}A'Wx$. If B is the "true" model for the mean squares from the populations, then we may obtain the contribution of the "true" parameters to these estimates by substituting B in the above expression i.e. from the rows of $(A'WA)^{-1}A'WB$. For our example, we obtain

$$\begin{pmatrix} E_1 \\ E_2 \end{pmatrix} = \begin{pmatrix} 1 & 0.25 \\ 0 & 0.25 \end{pmatrix} \begin{pmatrix} E_1 \\ D_R \end{pmatrix}$$

i.e. $\hat{E}_1 = E_1 + \frac{1}{4}D_R$

$\hat{E}_2 = \frac{1}{4} D_R$

The algebraic specification of these biases does not depend directly upon the magnitude of the "true" population parameters, but only on the sources of variation present, the experimental design and the weights. Obviously the numerical value of the biases does depend on the parameter values. However, there is only one set of biases for each model for each of the fourteen population types described earlier. These are present for all the possibilities we considered in the "A" tables of Appendix G.

SECTION 3: THE RESULTS

1 GENERAL COMMENTS

The analyses described in the previous section performed on the different "data sets" derived from the fourteen hypothetical populations have produced results which are far too extensive to discuss in all their details. These results are summarised in the tables in Appendix G and largely speak for themselves, so that the reader can draw his own conclusions from the particular tables most relevant to the problem he is likely to encounter when planning experiments designed to detect and estimate certain effects. In this discussion we shall confine ourselves to making points which illustrate the general usefulness of the Tables and to considering particular effects which have been the subject of much controversey in recent years i.e. genotype-environment interaction and genotype-environment covariation (e.g. Moran, 1973; Block and Dworkin, 1974a, 1974b; Layzer, 1974; Lewontin, 1974; Linn, 1974; Feldman and Lewontin, 1975; Plomin et al, 1977). We will attempt to answer critics who claim that the complexity of the causes of variation in human behaviour makes any experiments designed to elucidate these causes not only worthless, but also misleading. We will do this by integrating information from the different analyses described earlier in order to determine the likelihood of seriously misleading ourselves in a variety of different circumstances. We will also consider claims that the biases that occur usually lead to the overestimation of heritabilities.

2 POPULATIONS WITH ADDITIVE VARIATION

We considered four population types in which variation was only produced by additive effects:

1. E_1D_R 2. $E_1E_2D_R$ 3. E_1D_RM 4. $E_1E_2D_RM$

These populations contain various combinations of the contributions of specific environmental variation, between families environmental variation, additive genetical variation and additional variation produced by assortative mating.

For the very simplest population (containing E_1 and D_R), rejection of the null hypothesis of no genetical variation should not prove difficult unless the heritable component accounts for less than 50 per cent of the variation, when over 600 individuals will be required. If we were to mistakenly accept the E_1E_2 model as an adequate explanation of our data, E_1 and E_2 would be equally biased by $\frac{1}{4}D_R$. The contribution of D_R to E_1 and E_2 is always $\frac{1}{4}$, when we omit to fit D_R when it is a source of variation in a population.

Detection of D_R is generally not too difficult, but experimental sizes of about 600 are required to detect E_2 when it accounts for 25% of the variation and over 1000 are needed when its contribution falls to 5%. Thus, we might fail to detect E_2 as significant in a modest experiment and in such an experiment we would be unlikely to reject the E_1D_R model because over a thousand individuals are needed to reject this model when the true population contains $E_1E_2D_R$ and E_2 accounts for 25% of the total variation. Similarly, quite large numbers are needed to reject the E_1D_R and E_1D_RM models when the true causes of variation are $E_1E_2D_RM$. From Tables 1A and 2A we see that failure to fit E_2 when it is

in fact present leads to biases of all the other parameters we might fit. If we fail to detect significant E_2, we are most likely to take E_1D_R as the true explanation for variation. In this case, the contribution of E_2 to $\frac{1}{2}D_R$ is slightly greater than that to E_1, thus leading to over-estimation of the narrow heritability in this experimental design.

The number of individuals required to detect assortative mating, when $h^2_N = 0.9$ is quite small (i.e. about 300), but when the heritability drops to 0.5, over 5000 individuals are required. Obviously we may fail to detect significant assortative mating when it is present. Both E_2 and M contribute to the same statistics in studies of twins reared together and it is possible in such cases that estimates of their joint effect may be significant when fairly small experimental sizes are used.

For the E_1D_RM population, we might easily accept E_1D_R as an adequate explanation of the data since the number of individuals needed to reject this false hypothesis is over a thousand for $h^2_N = 0.9$ and over 12000 for $h^2_N = 0.5$. For the $E_1E_2D_RM$ population either E_1D_R or $E_1E_2D_R$ might be accepted depending upon the sample size. The contribution of M to E_1 and E_2 is very small and $\frac{1}{2}D_R$ is estimated approximately as $\frac{1}{2}D_R + \frac{1}{2}M$. However, the direction of the bias although small is such that in the case of the E_1D_R model, the true narrow heritability ($h^2_N = \frac{1}{2}D_R + \frac{1}{2}M/\ ^2_T$) is overestimated by the narrow heritability based on the false E_1D_R model. When the $E_1E_2D_R$ model is accepted for the $E_1E_2D_RM$ model, estimates of narrow heritability are not biased.

In these simple situations, narrow heritabilities may thus be overestimated, providing some support for the claims of critics. However, these same critics express the opinion that the causes of individual differences in human behaviour are not simple and so the situation for more complex populations must be examined.

3 POPULATIONS WITH DOMINANCE

The results for four populations are presented in Appendix G: $E_1D_RH_R$, $E_1E_2D_RH_R$, $E_1D_RH_RM$, $E_1E_2D_RH_RM$. We will discuss mainly the simplest of these.

Eaves (1972) showed that the detection of dominance using several experimental designs involved prohibitive sample sizes. We amply confirm his findings in each of the four types of population considered here. The smallest sample in which dominance could be detected in any of the populations was 4643 in the $E_1D_RH_RM$ population, when the broad heritability was 0.93 and the dominance ratio was 0.5. In most other cases the numbers required were much higher. Thus detection of dominance seems to be virtually impossible even with this "ideal" experimental twin design. In practice, in more restricted data sets, we will be extremely unlikely to detect dominance, so that the collection of twin data is likely to prove of little use for the study of genetic architecture (Mather, 1943) in human populations. Inbreeding studies or data on other types of relationships (e.g. half siblings, grandparents, etc.) are needed if we wish to detect dominance (see Jinks and Fulker, 1970).

When E_1, D_R and H_R are the true causes of variation in the population, the "false" E_1D_R model is most easily rejected when $h^2_N = 0.90$ and the dominance ratio is 1.0. Even in this case 16042 individuals are required before rejection at the 5 per cent level will occur in 95% of cases, and the false E_1D_R model will be accepted as representing the true causes of variation.

From Table 3A, we find that:

$$\hat{E}_1 = E_1 + 0.02\ H_R$$

$$\hat{D}_R = D_R + 0.47\ H_R$$

In other words, estimates of E_1 are not seriously biased and H_R contributes to the estimate of additive genetical variation. Estimates of the narrow heritability obtained by fitting the E_1D_R model include both additive and dominance variation and give us the broad heritability for the trait.

If we now consider an additional source of environmental variation, so that the true population contains E_1, E_2, D_R and H_R, we find that the number of individuals required to reject the E_1D_R model drops considerably to about 1100, when $h^2_N = 0.9$ and 2100, when $h^2_N = 0.5$. This number is still fairly large and in many typical experiments, we would accept the E_1D_R model as the true explanation of the observed variation.

If we did this, we would estimate:

$$\hat{E}_1 = E_1 + 0.46\ E_2 + 0.01\ H_R$$
$$\hat{D}_R = 1.05\ E_2 + D_R + 0.49\ H_R$$

H_R again only contributes to $\hat{D}_R$, but our estimates of $\hat{E}_1$ and $\hat{D}_R$ are biased by the E_2 component which we have also failed to detect. Estimates of h^2_N are thus inflated by the E_2 component, as we saw in the last section.

If we had performed an experiment of well over 2500 individuals, we would probably reject the E_1D_R model. In this case we would then fit a series of three parameter models: $E_1E_2D_R$, $E_1D_RH_R$, E_1D_RM. From Table 4B.1, we can see that the $E_1D_RH_R$ model would almost certainly be rejected but that either the $E_1E_2D_R$ or E_1D_RM models would probably be retained as possible models for variation.

The contribution of the two parameters to the estimates for the $E_1E_2D_R$ model is:

$$\hat{E}_1 = E_1 + 0.01\, H_R$$
$$\hat{E}_2 = E_2 - 0.01\, H_R$$
$$\hat{D}_R = D_R + 0.51\, H_R \simeq 2\, V_G,$$

where V_G is all the genetic variation.

For the $E_1 D_R M$ model it is:

$$\hat{E}_1 = E_1 + 0.51\, E_2$$
$$\hat{D}_R = D_R + 0.55\, E_2 + 0.67\, H_R$$
$$\hat{M} = 1.38\, E_2 - 0.16\, H_R$$

Thus, the $E_1 E_2 D_R$ model would lead us to estimate the narrow heritability as the proportion of variation produced by all genetical effects i.e. the broad heritability and we would not underestimate the importance of environmental effects.

In the case of the $E_1 D_R M$ model, the narrow heritability will be inflated, but the importance of the environmental components will be overestimated.

Our conclusion then is that in populations in which either E_1, D_R and H_R or E_1, E_2, D_R and H_R are the true causes of variation, we will not overestimate the importance of genetical effects except when we are unable to reject the $E_1 D_R$ model for the $E_1 E_2 D_R H_R$ population because the power of the test for E_2 is not high enough. Samples of over a thousand will be needed to reject this model when $h^2_N \simeq 0.5$ and of over two thousand when $h^2_N \simeq 0.9$.

Similar conclusions emerge when we consider additional variation produced by assortative mating. When E_1, D_R, H_R and M are the true causes of variation, it is difficult to reject the $E_1 D_R$ model and we may estimate:

$$\hat{E}_1 = E_1 + 0.01\, H_R - 0.05M$$
$$\hat{D}_R = D_R + 0.48\, H_R + 1.18\, M \simeq 2V_G$$

Again, all genetical effects are estimated largely with $\hat{D}_R$ and estimates of narrow heritability reflect all variation produced by genetical differences between individuals.

In the population where $E_1E_2D_RH_R$ and M produce variation over 1000 individuals are needed for rejection of the E_1D_R model. If we fail to reject this model, parameter estimates will be biased by E_2 and the proportion of variation attributable to environmental effects will be underestimated. If E_2 is fitted, then $\hat{E}_1$ and $\hat{E}_2$ are virtually unbiased and $\hat{D}_R$ contains all the genetical variation produced by D_R, H_R and M.

In the last two sections we have seen that failure to detect H_R does not introduce biases of the sort that lead us to underestimate the importance of environmental factors. Failure to detect E_2 is also likely with typical sample sizes, and this may lead to overestimation of the importance of genetical factors.

4 GENOTYPE-ENVIRONMENT INTERACTION

We have generated four populations in which there is variation produced by genotype-environment interaction, where additive genetical effects interact with the common family environmental component:

1. $E_1E_2D_RI$
2. $E_1E_2D_RI\ M$
3. $E_1E_2D_RH_RI$

From Tables 7B to 10B, we can see that the number of individuals needed to detect I as significant at the 5 per cent level in 95 per cent of similar experiments is prohibitive. The number needed to detect the E_2 component is also large. The population where we are most likely to detect GXE is one in which the true causes of variation are E_1, E_2, D_R and I and $h^2_N = 0.9$ and GXE accounts for twenty per cent of the total variation. Even in this case 2958 individuals are needed. Thus, GXE which is not systematically related to genetical or environmental deviations, and is, therefore, not detectable by other methods, is unlikely to be detected by fitting linear models to mean squares unless the contribution of GXE to the total variation is much larger than 20 per cent. Work in other organisms suggests that GXE is not likely to be much larger than this. There is thus some evidence that GXE may remain undetected in man, as claimed by several authors. We may now ask whether this will lead to serious biases of parameter estimates from models found adequate to account for variation using the chi-square criterion.

The number of individuals needed to reject the E_1D_R model is moderate in this instance:

h^2_B	I	N
0.5	10%	802
0.5	20%	636
0.9	10%	689
0.9	20%	479

When the heritability is high and GXE accounts for 20 per cent of the variation, only a relatively small experiment is required. But when the heritability is lower and GXE accounts less than 10% of the variation, then approaching 1000 individuals may be needed. If our experimental size is such that we mistakenly accept the E_1D_R model, then we estimate:

$$\hat{E}_1 = E_1 + 0.45\ E_2 + 0.25\ I$$
$$\hat{D}_R = D_R + 1.07\ E_2 + 0.25\ I$$

The GXE contributes equally to E_1 and D_R, which means that it biases the contribution of E_1 to the total variation twice as much as that of D_R, since the contribution of D_R to the total variation is only $\frac{1}{2}$. This would lead to the underestimation of h^2_N but for the fact that we have failed to fit E_2 and as before this leads to the over-estimation of h^2_N.

If we have a reasonable experimental size, we are likely to reject the E_1D_R model. We would then fit three 3 parameter models: $E_1E_2D_R$, E_1D_RM and $E_1D_RH_R$. Rejection of the $E_1D_RH_R$ model would be immediate, not only because few individuals are required, but also because estimates of H_R would be non-significant and negative. However, both the $E_1E_2D_R$ and the E_1D_RM model would fit the data, leading to estimates:

1. $E_1E_2D_R$ model

$$\hat{E}_1 = E_1 + 0.07\ I$$
$$\hat{E}_2 = E_2 + 0.25\ I$$
$$\hat{D}_R = D_R + 0.43\ I$$

2. E_1D_RM model

$$\hat{E}_1 = E_1 + 0.51\,E_2 + 0.25\,I$$
$$\hat{D}_R = D_R - 0.53\,E_2 + 0.50\,I$$
$$\hat{M} = 1.36\,E_2$$

In the latter case, all estimates are biased and our failure to fit E_2 leads to overestimate of the narrow heritability.

When we fit $E_1E_2D_R$ to the "data", $\hat{E}_1$ is virtually unbiased. However, both $\hat{E}_2$ and $\hat{D}_R$ are quite strongly biased by the GXE, when its effect is ignored in fitting the model. Since the contributions of additive genetical variation and cultural effects to the total variation are $\frac{1}{2}D_R$ and E_1, these effects are almost equally biased by I.

We find that estimates of the narrow heritability calculated on the assumption of no GXE are inflated, and that we are likely to overestimate the true contribution of additive genetical variance, when the estimate is based on the relatives in our experimental design.

This provides some justification for the criticisms of certain authors (e.g. Moran, 1973; Layzer, 1974; Feldman and Lewontin, 1975) that unsystematic GXE, if it is an important source of variation in human populations may remain undetected and inflate heritability estimates. However, this will not be the case for all experimental designs. GXE could lead to the underestimation of D_R in transgenerational data (e.g. parent-offspring covariances).

We may be interested in analysing variation in human behaviour for two main reasons. The first is a desire to discover the biologically important sources of variation in human behaviour and the second a desire for environmental intervention to enhance the phenotypes of those whose genetic predisposition is for a lower phenotype of some particularly desirable attribute. In either case, we should not exaggerate the

significance of our findings about the difficulties that face those who attempt to detect non-systematic GXE. The existence of GXE systematically related to genetical or environmental deviations is likely to be far more meaningful biologically and could suggest suitable medical or educational strategies for the enhancement of particular phenotypes. Several tests have been proposed for the detection of such systematic GXE (Jinks and Fulker, 1970) and these are discussed in earlier parts of this thesis.

5 GENOTYPE-ENVIROMENT COVARIANCE

5.1 Cultural Transmission

In this population, genotype-environment covariation arises as a result of the cultural impact of parents on their offspring. The specification of this population was described in Section 1.2.13. Since only one new parameter, b, was involved in the specification of E_2 and Cov g_2e_2, we could have estimated E_1, D_R and b by the more powerful approach of non-linear weighted least squares (Eaves, 1975). However, the approach described in Section 2 was continued and four parameters were estimated: E_1, D_R, E_2, C, where C is the genotype-environment covariance produced by cultural transmission. The constraint upon the values of E_2 and Cov g_2e_2 implied by the cultural transmission model was, therefore, not enforced, but since real data may not exactly follow our model, this may be a more useful approach.

From Table 10B.2, we can see that the detection of variation produced by cultural transmission is not likely to be so difficult as for some of the effects considered previously. When b = 0.25, the number of individuals required to detect C was 631 and 461, for the two levels of heritability. Detection of E_2 was also feasible (635 and 249 individuals). When b = 0.1, larger (but not unreasonable) samples are required for C (2000 and 1231) although the detection of E_2 was not so likely (25332 and 1521). Sample sizes, of course, increase as the heritability increases and there is a corresponding decrease in the size of E_2 and C.

Another heartening feature of Tables 10B.1 and 10B.2 is that the number of individuals needed to reject false models is generally lower

than that seen in previous Tables. This suggests that we will be less likely to mislead ourselves about the true causes of variation when cultural transmission is present.

We can obtain the linear combination of the observed statistics which yields the weighted least squares estimate of 2 Cov g_2e_2. The coefficients of the eleven statistics in the estimator of 2 Cov g_2e_2 for the case when b = 0.25 and $\frac{1}{2}D_R = 2E_1 = 112.5$ are:

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

This shows that the test for Cov g_2e_2 depends largely upon a comparison of the total variances of individuals reared by their natural parents with that of individuals reared by foster parents. It does not matter whether the individuals are twins or not, or even whether they are reared in pairs. Thus, information about genotype-environment covariance produced as a result of cultural transmission would be best sought in adoption studies, as suggested by Jinks and Fulker, (1970) and Cavall-Sforza and Feldman (1973). The use of such adoption studies

for the detection and estimation of the parameters of cultural transmission was discussed extensively in Part B: A Twin Study of social attitudes.

There has been discussion in the literature about the contribution of dominance and genotype-environment covariation to variation in intelligence. We fitted several models incorporating dominance to the four data sets generated from the population where E_1, E_2, D_R and C were contributing to variation:

1. $E_1D_RH_R$
2. $E_1D_RH_RM$
3. $E_1E_2D_RH_R$

In all cases, estimates of dominance were small and negative. The $E_1E_2D_RH_R$ population is the most similar to the "true population" containing $E_1E_2D_RC$. Even in the situation most favourable for the detection of the false H_R effect, 1417 individuals were needed for H_R to be significant in this experimental design. When b = 0.25, fairly small sample sizes only are required for the false $E_1E_2D_RH_R$ model to be rejected at the 5 per cent level in 95 per cent of cases, although rather more were needed when b = 0.1.

The likely failure of the $E_1E_2D_RH_R$ model and the negative values obtained for dominance variation suggests that we would not accept the $E_1E_2D_RH_R$ model, but would seek an alternative explanation for the pattern of variation. It may, however, be possible to mistake negative genotype-environmental covariance with dominance in borderline cases, because some of the consequences of negative CovGE may reproduce certain of the effects of dominance. This may be true for Cov GE produced both by cultural transmission and also by sibling effects

(in the case of competition). We have specified populations in which competition is operating (see Section 1.2.15 and 1.2.16) and so we can study the likelihood of being mislead about the true causes of variation in such cases.

5.2 Sibling Effects

A second source of genotype-environment covariation was defined by Eaves (1976a) in which the Cov GE arises because the phenotype of a sibling produces an environmental effect on the phenotype of his co-sibling. An advantageous phenotype in one sibling may enhance the environment necessary for a co-sibling to develop high levels of the advantageous trait (i.e. cooperation), or an advantageous phenotype may develop in one sibling at the expense of a co-sibling (competition). These lead to positive and negative genotype-environment covariance respectively.

Power calculations and estimation of biases were carried out for two populations: $E_1D_RD_R''D_R'$ and $E_1E_2D_RD_R''D_R'$ and the results are presented in Tables 11A.1 to 12B.15. The discussion will be confined mainly to a consideration of the first population type, of which we considered sixteen different cases with the relative contributions of the parameters varying according to h^2_N, the direction of the covariance, the relative magnitude of D_R'' compared with D_R and the value of f_{ge}.

The correct $E_1D_RD_R''D_R'$ model was fitted to each data set and the number of individuals needed to detect the four sources of variation was determined as described earlier. From Table 11B.6, it can be seen

that no more than 142 individuals are needed to detect E_1 in an experiment of this design and no more than 327 are needed to detect D_R. The numbers needed to detect D_R'' and D_R' vary considerably depending on the values of the population parameters. If we average the numbers required over values of ρ_{ge} and relative magnitudes of D_R'' compared tieh D_R', then when $h^2_N = 0.5$, 1969 individuals are required to detect D_R'' and only 513 for D_R'. When $h^2_N = 0.9$, the numbers are 989 and 244.

Thus generally we notice that it is easier to detect D_R' and D_R'' as the heritability increases. It is considerably easier when the environmental effect of the alleles equals their direct effect upon the genotype i.e. when $D_R'' = D_R'$. If the proportion of loci contributing to both D_R and D_R' is one, the likelihood of detecting D_R' is considerably increased at the expense of the power of the test for detecting D_R''. The power of the test for detecting sibling effects is greater when these covary with the direct effects of genetical differences (i.e. when ρ_{ge} is positive). The fact that far fewer individuals are needed to detect D_R' than D_R'' arises because the test of D_R'' depends merely on a comparison between the total variances of individuals reared as pairs and singletons, whereas the significance of D_R' depends upon a very precise pattern of expectations which involve the individual mean squares as well as the total variances.

It seems that the detection of both competition and cooperation should be relatively easy in the majority of non-extreme situations which we have considered. In order to emphasize this further we may list the other models which might be fitted to mean squares derived from populations in which E_1, D_R, D_R' and D_R'' are the true causes of variation and show why we might reject those in most cases.

1. E_1E_2 - significant residuals (and estimates of $\hat{E}_2$ non-significant or negative when D_R' negative)
2. E_1D_R - residuals significant in most situations
3. $E_1E_2D_R$ - significant residuals or negative estimates of $\hat{E}_2$ or non-significant estimates of $\hat{E}_2$
4. $E_1D_RH_R$ - significant residuals, estimates of $\hat{D}_R$ or $\hat{H}_R$ negative and non-significant, depending on whether ρ_{ge} is positive or negative
5. E_1D_RM - significant residuals or M sometimes non significant (or negative when ρ_{ge} negative)
6. $E_1E_2D_RH_R$ - significant residuals and negative or non-significant estimates of $\hat{D}_R$ or $\hat{H}_R$ or $\hat{E}_2$ depending on the particular situation.
7. $E_1E_2D_RI$ - negative estimates of $\hat{E}_2$ or $\hat{I}$
8. $E_1E_2D_RM$ - estimates of $\hat{E}_2$ and $\hat{M}$ non-significant or negative, residuals significant
9. $E_1D_RH_RM$ - $\hat{D}_R$ or $\hat{M}$ negative. $\hat{D}_R$, $\hat{H}_R$ and $\hat{M}$ often non-significant. Residuals often significant

In all these cases, if we fitted a "false" model to the "data", we would find some reason for rejecting the false model in this experimental design. Thus, it is not easy to be mislead about the true causes of variation in a population when these are E, D_R, D_R'' and D_R'.

The cases of the $E_1D_RH_R$ and $E_1E_2D_RH_R$ models are particularly pertinent in view of the discussion of the roles of dominance and Cov GE in the determination of I.Q. We suspected that it might be possible to confuse these two effects when Cov GE is negative. However, from the information provided in Tables 11B.4 and 11B.7, it seems that

this would be unlikely. Overall, we have every reason to be optimistic about our chances of detecting Cov GE produced by sibling effects and rejecting alternative false models.

SECTION 4: DISCUSSION

In this work, we have attempted to provide precise statements about the power of tests for detecting components of variation and for rejecting false models. Biases in estimated parameters can be calculated mathematically given that we simulate a series of possible populations, from which mean squares can be derived for a particular experimental design. Thus, we have quantified many of the arguments which are discussed, without evidence, in the field of human behaviour genetics. We have done this for two main reasons:

1. In order to provide a rational basis for the discussion of criticisms of the analysis of individual differences in human behaviour.
2. In order to provide a quantitative framework for the planning of future research.

We have shown that in some cases the critics may have a case, although they had no evidence for this themselves. Consider, for example, the case of GXE, where we have shown that its effects may remain undetected even in extensive twin studies and can bias estimates of the components of variation, leading to an overestimate of the importance of genetical differences between individuals. This provides some support for authors such as Moran, (1973), Layzer (1974a) and Feldman and Lewontin (1975). However, it does not support the claim that GXE precludes any worthwhile analysis of individual differences in human populations. Non-systematic genotype-environment interaction may be detected with more appropriate experimental designs, and in any case systematic GXE is likely to be more biologically and socially relevant. This source of variation can be detected using several scaling tests (Jinks and Fulker, 1970).

In contrast to this, we have found that the detection of Cov GE produced by either cultural transmission or sibling effects should prove relatively easy. Much of the misunderstanding and controversy surrounding non-additive effects such as GXE and Cov GE arises because previously the specification of these effects was purely empirical. However, the work of Eaves (1976a, 1976b) has provided a theoretical framework from which a knowledge of the consequences of Cov GE in particular populations can be developed through simulation and practical studies.

These simulations have provided a wealth of information about the sorts of experiments that should be conducted in order to test certain hypotheses about the causes of individual differences, mainly in terms of the sample sizes required and the biases that may be introduced when false models of variation are retained because of an inadequate experimental design.

The disadvantage of such simulations is that while they may tell us a lot about what we may expect given one particular constellation of relatives, they cannot be generalised to other experimental designs. Other types of relatives or studies may detect effects which cannot be detected in this study. The precise nature of the biases and the powers of particular tests needs to be determined for each specific situation.

This leads us to suggest a future development of the present work. A simulation package of computer programs should be developed for use by psychogeneticists. This package would produce any desired population type, and calculate the mean squares derived from this population for specified experimental designs, and then perform the

calculations we have described in Section 2. Thus before embarking on a study designed to test a specific hypothesis, a worker could determine the optimal experimental design within his resources and would know beforehand the probability of detecting particular effects and the biases he could expect. In this way many fruitless research programmes would not be started, and hopefully the concentration of resources on a few efficient experiments would do much to increase our understanding of the nature of individual differences in human behaviour.

GENERAL SUMMARY AND DISCUSSION

This discussion will be brief since each part of the thesis is largely self-contained and has already been discussed in some detail. The general significance of the findings and their implications for future research will be considered.

Human behaviour genetics uses the methods of modern genetics to study traits which were once considered solely as the interest of psychology and shows the importance of the biological basis of behaviour for understanding the evolution of culture and a social system in man. The relationship between biology and behaviour can only be understood by studying the evolution of behavioural traits through their contribution to fitness and the dependence of an organism on and its sensitivity to the environment. The determination of genetic architecture and understanding of genotype-environment interaction and the non-genetic, as well as the genetic, transfer of information from generation to generation are essential. The analysis of a few twin correlations and unrealiable estimation heritabilities are clearly inadequate for this purpose, although in the past most studies were confined to these simple aims. The limitations of this approach, suggested by Jinks and Fulker (1970), were illustrated using the data from the Georgia Twin Study and the advantages of the analysis of variance components was shown. Whites were more variable than Blacks, a fact which would have remained undiscovered in an analysis of correlations, leading to biased estimates of genetical and environmental effects and probable errors of inference about the causes of individual differences in ability. The greater variability of Whites could be partly attributed to greater between families differences, due to cultural differences

within populations or the greater genetical heterogeneity of Whites. Some of this heterogeneity could be attributed to differences in mean performance between Whites from different regions. The roles of the mating system and culture in maintaining differences between sub-populations and Whites in the United States could provide a constructive area for future investigation.

Although the analysis of variance components represents an advance over the analysis of correlations, the test battery and the size and design of the experiment proved limiting factors in the Georgia Twin Study. High levels of test specific variation were found for certain Spatial tests, and produced many ambiguities during the analyses. This suggests that great care should be taken at the initiation of an enquiry in choosing reliable, standardised tests. More work is obviously needed to determine the tests most suitable for use with White and Black populations.

The unreliability of certain tests, leads to inflated within families variation in MZ twins and reduced MZ twin correlations. This made discrimination between simple genetical and simple environmental models difficult with the sample sizes used here. Power calculations showed that the number of twins required to reject one of these hypotheses at the 5% level in 95% of cases is greater than the number in this study, when the MZ twin correlation is about 0.5 - 0.6. A larger twin study of Blacks and Whites in the US using a more carefully chosen test battery would enable the causes of variation in ability, particularly Spatial ability, to be determined more satisfactorily.

The twin study of social attitudes was large enough to test the genotype-environmental models that can be specified for twins. Estimates

of additive genetical variation, specific environmental variation and a between families component were remarkably consistent with those from two previous studies. However, the resolution of the between families component into variation produced by cultural differences and the mating system is not possible with data only on twins reared together. Twin studies can be useful in the early stages of an investigation for providing powerful tests of certain simple hypotheses, but do not permit the detailed investigation of sources of variation. Thus, there is a general trend towards the collection of other types of data, and particularly data on whole families. In Birmingham and London, a study of ability, personality and attitudes is attempting to combine the features of twin and family data. Mothers and fathers are being tested, as well as the twins themselves in an extensive programme. Thus, many additional hypotheses can be tested concerning, for example, the causes of variation in adults compared with children and sex-linkage. Nance and Corey (1976) and Eaves (personal communication) have suggested an improved design which allows the detection and estimation of most effects of interest. This includes measurements on pairs of adult identical and non-identical twins, their spouses and offspring and thus provides covariances between twins, spouses and parents and offspring. Relationships among the children of a pair of MZ twins provide the equivalent of half sibling and sibling covariances. Similarly the offspring of DZ twins provide sibling and cousin covariances. However, for the resolution of cultural differences, adoption data provide powerful tests (although the equivalent of covariances between parent and offspring (not reared by him) is provided by an MZ twin and the offspring of his co-twin). Adoption data were used to investigate the role of cultural transmission for social attitudes.

Both twin and adoption data are open to the criticism that the individuals may be atypical of the population as a whole The collection of data on a random sample of families as they naturally occur in the population would provide sufficient numbers of most types of relationships if large enough samples were employed and would overcome the problem of individuals being atypical. However, the unbalanced pedigrees obtained could make analysis difficult and wasteful of information. The method of maximum likelihood estimation by pedigree analysis, used for analysing the adoption data of the social attitudes study shows how this problem may be overcome. In this case there were too few individuals of some types e.g. natural-adopted sibling pairs, which meant that tests of hypotheses of particular interest were not adequate. There was no evidence for the significant effect of cultural transmission and the between families component could apparently be explained by reference to the mating system in these data. However, the power of the test was probably low.

In Part C, methods were introduced for estimating the power of the test for detecting particular effects, the power for rejecting false hypotheses and for calculating biases when false models are fitted. It was shown that large samples are needed for the detection of genotype-environment interaction, but not for genotype-environment covariance. The exclusion of CovGE from models of variation leads to failure of alternative models with relatively small sample sizes, and is unlikely to lead to overestimation of the genetical component. This is not true when GXE is mistakenly omitted from analyses, but the biases are not necessarily large.

The limitation of all such simulations is their specificity to

particular experimental designs, populations and actual models fitted. However, the methods may be generalised and used to estimate biases and determine probable errors of inference in other situations. One aim of future work should, therefore, be to construct a general simulation package which enables the possibilities for any experimental design, population and model to be tested. This would be useful in determining the errors of inference that might have occurred when interpreting results. However, its main value would be in the optimisation of experimental designs for the detection and estimation of effects of interest before a study is undertaken. This should prevent waste of time and resources on inefficient studies which in the end cannot make powerful tests of the hypotheses they were designed to test.

Many of the problems of power and unreliability in the Georgia Twin Study were overcome by extracting and analysing a "general factor" of ability. A factor analysis of the data showed that the tests could be regarded as belonging either to a Verbal Arithmetic factor or to a Spatial factor. In order to consolidate the conclusions from this study, scores of the two factors should be obtained and analysed. An alternative approach would be to analyse the structure of the variance and covariance between tests into its genetical and environment components, using method already available (Joreskog, 1973; Martin and Eaves, 1977).

However, the main development of the work in this thesis should be in the analysis of the environment. An analysis of "genotype-environment interaction" showed that, for ability, many apparent interactions between genotypic and environmental differences may be regarded simply as a property of the scale of measurement rather than the

individuals sensitivity to his environment. Little evidence for "GXE interactions" could be found for social attitudes. Thus, the analysis of cultural differences and cultural transmission is our main problem in relating the environment to the phenotype of an organism. When the environment is provided by the phenotype of the parents, cultural transmission may arise. The work in this thesis suggests further study of the effects of culture in the evolution of man's behaviour using studies appropriately designed with the aid of computer simulations. Simulation studies could also be used to study the consequences of cultural transmission and assortative mating for the pattern of variation. Alternatively, family data could be collected and analysed using the methods of pedigree analysis. In this thesis the problem of the joint role of assortative mating and cultural differences has been resolved in terms of path models. Future work should include the development of Fisher's (1918) model of assortative mating to include cultural transmission as specified by Eave's (1976b) model. This would enable further advances to be made in understanding one of the most interesting current problems in human behavioural genetics - the evolution of the behaviour leading to cultural differences and cultural change and the development of society.

APPENDIX A: FINAL DETERMINATION OF ZYGOSITY FOR ELEVEN PAIRS OF TWINS OF UNCERTAIN ZYGOSITY

Twin Pair No.233 - These 16 year old white girls were called MZ by the discriminant analysis program and DZ by AID. The girls were exactly the same height but different by 14.5% in weight. Differences in head length and breadth were also significant. One sister was right-handed; the other ambidextrous. The twins reported they were rarely misidentified. They believe they are DZ. Twin A says, "There is no resemblance. Everything is unlike." Final classification, DZ.

Twin Pair No.277 - In terms of biometric measurements, these 14 year old Negro girls appear to be identical. They are the same height. Head length and head breadth are also the same. There are only slight differences in the other physical measurements. However, Twin A is right-handed; Twin B is not. The test for colour blindness probably convinced the investigator. Final classification, DZ.

Twin Pair No.282 - These 14 year old Negro boys were classified MZ by AID and DZ by the discriminant analysis program. Weight difference was 6%; face length difference, 8%. Twin A was colour blind; Twin B, not. A is left-handed; B is right-handed. Final classification, DZ.

Twin Pair No.284 - These 15 year old Negro girls were not classified the same way by the computer programs. Examination of their files convinced the investigator they were DZ. The twins say they are fraternal. A is left-handed; B is right. Both say they don't look alike. Both twins say that Twin A is darker skinned and heavier. But also both say that their nose, mouth and eyes look alike. Teachers, parents, and friends sometimes mistake one for the other. Differences in head length and breadth both are significant at the .01 level from Verschuer tables. Final classification, DZ.

Twin Pair No.309 - This pair of 17 year old white girls was classified DZ by AID and MZ by discriminant analysis. The girls differ by 8% in height

and 27% in weight. Twin A is right-handed; B is left-handed. The attending physician said they were DZ and the girls believe they are fraternal. Final classification DZ.

Twin Pair No.317 - These are 14 year old white girls. A says she knows she is an MZ twin; B is just as confident she is DZ because the attending physician said they were DZ. In the questionnaire B said their noses were not alike. This, in fact, is the case since their noses differ in length by 9%. Height difference is 5%; weight 13%. A is right-handed; B is left-handed. Final classification, DZ.

Twin Pair No.347 - This pair of 14 year old white girls says their attending physician said they are identical. However, Twin B says, "We look nothing alike." A's hair is brown; B's auburn. They never, or only rarely, are mistaken by teachers and parents. Differences in nose length, face length, head length, and height all support final diagnosis of DZ.

Twin Pair No.362 - These 13 year old Negro boys "know they are identical." But Twin A says that B's hair grows faster than his. They are only occasionally mistaken by teachers, friends and parents. Differences in five biometric measurements, height, weight, head breadth, nose length and face length convinced the investigator of the final DZ classification.

Twin Pair No.373 - These 13 year old Negro girls know they are fraternal. A is right-handed; B is left-handed. They are rarely mistaken by friends, teachers or parents. A's hair is lighter and thinner than B's. Both twins report their face, legs, and head to be different. The AID program called them DZ; the discriminant analysis MZ. Final classification DZ.

Twin Pair No.375 - These 17 year old white girls say they are identical but rarely misidentified. They indicate their nose, fingers, hands, stomach and bust to be similar. The discriminant analysis program classifies the girls as DZ, the AID program as MZ. Rorhrer's Index of Body Structure and Kaup's Index both support the diagnosis of DZ. Differences in nose length and face length confirm the final DZ classification.

Twin Pair No.379 - These 16 year old Negro boys say they look alike and know they are identical because their attending Physician said they were. They are seldom misidentified. Face length difference is the only biometric measure that supports a DZ diagnosis. Other measurements are within MZ limits. Final classification MZ.

APPENDIX B: PSYCHOLOGICAL TESTS

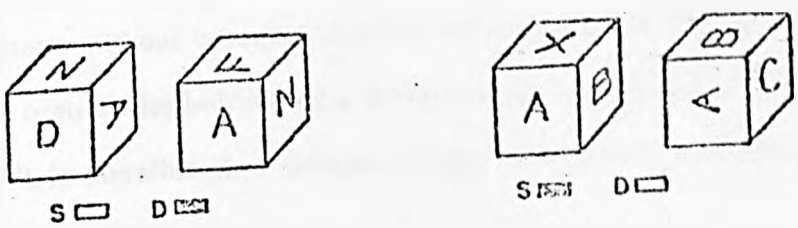
CALENDAR TEST (Calendar)

On the Calendar Test, developed by Remondino (1962), the examinee is asked to check the exactness of 50 sentences pertaining to the relationship of the days of the week. In a factor analysis, Remondino found that this test loaded on the Number factor. Following are examples of the type of questions asked:

If today is Sunday then tomorrow will be Monday. T F
If yesterday was Wednesday then today is Saturday. T F

CUBE COMPARISONS TEST (Cubes)

The Cube Comparisons Test was developed from Thurstone's Cubes. Each item presents two drawings of a cube, such as shown below. Assuming no cube can have two faces alike, the subject has to decide whether the two drawings can represent the same cube or must represent different cubes. The instructions indicate that the task can be performed (1) by mentally turning one of the cubes so that the face of one cube is oriented in the same way as the like face of the second cube and then comparing the sides one by one or (2) by noting whether two faces which are side by side have the same letters or numbers in the same position relative to one another. The process of obtaining the answers by the second method consists largely of verbal reasoning although it does require a "static" awareness of three-dimensional relations as opposed to a more "dynamic" moving around of the blocks in space.



SIMPLE ARITHMETIC TEST (Arithmetic)

The Simple Arithmetic Test, taken from an unpublished study by Mukherjie (1963), contains seven parts, each consisting of a number of simple arithmetical problems. Part 1 contains 15 problems; part 2, 20 problems; and parts 3 through 7 each contains 25 problems. Speed is an important factor in this test since the examinee is allowed only two minutes per test. This is a multiple-choice test with five alternatives for each problem. The complexity of the problems decreases from part 1 to part 7. Examples contained in each part are given below:

Part 1: $\frac{4(77+39-4)}{7}$ = 60 68 74 64 **84**

Part 2: 5(69+18-3) = 420 400 410 415 425
Part 3: 69+25-9 = 85 95 90 89 80
Part 4: 640 ÷ 5 = 120 128 88 136 126
Part 5: 8 x 91 = 738 728 732 739 737
Part 6: 19 - 7 = 12 13 14 15 16
Part 7: 83 + 17 = 90 110 100 109 101

WIDE RANGE VOCABULARY TEST (Vocabulary)

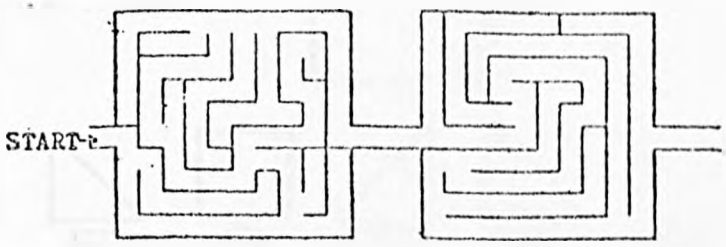
The Wide Range Vocabulary Test, which was adapted from a Cooperative Vocabulary Test (ETS; French, Ekstrom, & Price, 1963) is a five-choice synonym test having items ranging from very easy to very difficult. Samples of the items follow:

JOVIAL: 1. refreshing 2. scare 3. thickset 4. wise 5. jolly
DULLARD: 1. peon 2. duck 3. braggart 4. thief 5. dunce

THE MAZES TEST (Mazes)

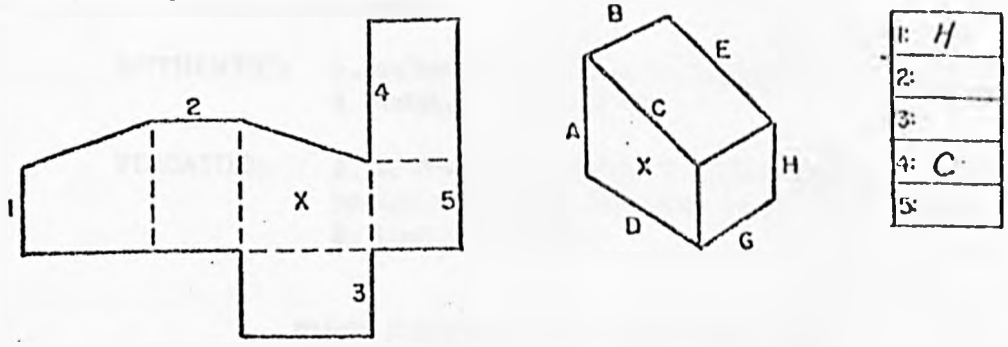
The Mazes Test was taken from a laboratory manual by McKinnon and Henle. The task, typical of earlier Maze problems, is to draw a line from one end to the other of the maze without crossing any line or entering blind alleys. Although this test does not require the holding of a 3-dimensional or even of a 2-dimensional figure in mind, it is possible that solution of the task will be facilitated by the ability

to remember briefly sections of the correct path before one draws a line. A sample maze is shown below.



SURFACE DEVELOPMENT TEST (Surface development)

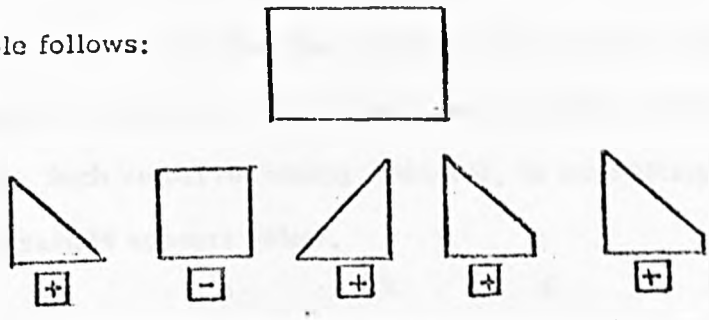
The Surface Development Test is adapted from Thurstone. In this test the subject has to imagine or visualize how a piece of paper can be folded to form some kind of object. Each item consists of a drawing of a piece of paper which can be folded on the dotted lines to form the object drawn at the right. (See sample below.) The subject is to imagine the folding, to figure out which of the lettered edges on the object are the same as the numbered edges on the piece of paper at the left, and to identify the letters of the answers in the numbered spaces at the far right. He is told that the side of the flat piece marked with the X will always be the same as the side of the object marked with the X. It appears that this task does require mental movement of the parts of the pattern and it is not likely that the subject can perform this task by verbal reasoning only.



FORM BOARD TEST (Form board)

Each item of the Form Board Test presents five shaded drawings of pieces some or all of which can be put together to form a figure presented in outline form.

The task is to indicate which of the pieces when fitted together will form the outline. An example follows:



SELF-JUDGING VOCABUIARY TEST (Heim)

The Self-Judging Vocabulary Test, developed by Heim et al (1965), contains two parts. The first part contains 128 words each of which the examinee marks with an A, B, or C (A = I know this word and could explain it to someone unfamiliar with it, B = I am doubtful as to what this word means, C = I have never seen this word before and have no idea what it means.) The second part of the test consists of the first 80 words of the 128 word list presented as a multiple-choice test with six alternatives. The second part of the test combines the advantages of the multiple-choice and creative answer techniques by allowing the examinee who thinks he knows the word but dislikes the six alternatives offered to write his answer in his own words below the six choices. In this study only the second part of the test is used. Examples of test items are given below:

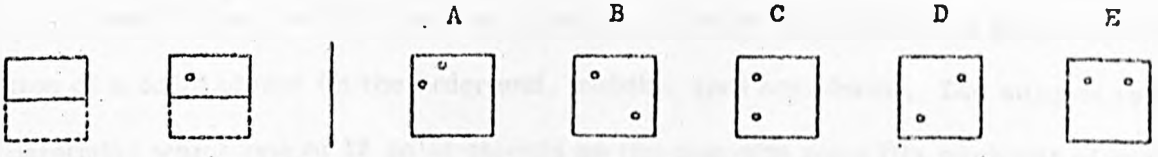
AUTHENTIC: 1. writer 2. to allow 3. respectful 4. a bargain 5. antique 6. genuine

VERSATILE: 1. of varied activities 2. pouring out 3. form of poetry 4. having masculine vigor 5. intense 6. kind of turnstile

PAPER FOLDING TEST (Paper folding)

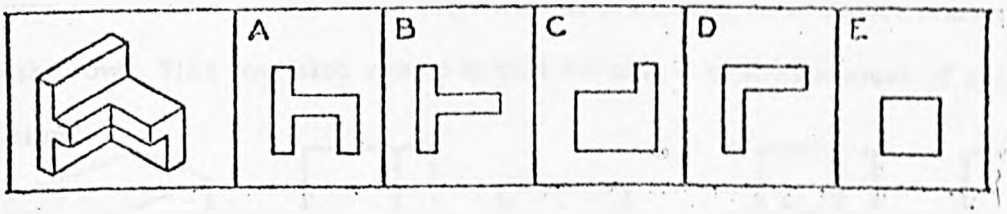
The Paper Folding Test was suggested by Thurstone's Punched Holes. For each item, successive drawings illustrate two or three folds made in a square sheet of paper. A drawing of the folded paper shows where a hole is punched in it. The

subject selects one of five drawings to show how the sheet would appear completely unfolded. While it is probable that the problems can be solved more quickly if one can easily imagine the folding and unfolding, one can also solve these problems by verbal reasoning. Such verbal reasoning, however, is more likely to lead to incorrect answers. A sample appears below.



OBJECT APERTURE TEST (Object aperture)

This test of spatial visualization, developed by Philip H. DuBois and Goldine C. Gleser, consists of a number of items similar to the sample below. A three-dimensional object is shown at the left, followed by outlines of five apertures or openings. The subject is to imagine how the object looks from all directions; then to select from the five apertures outlined the opening through which the solid object would pass directly if the proper side were inserted first. This usually requires the subject to mentally turn the object into other positions.



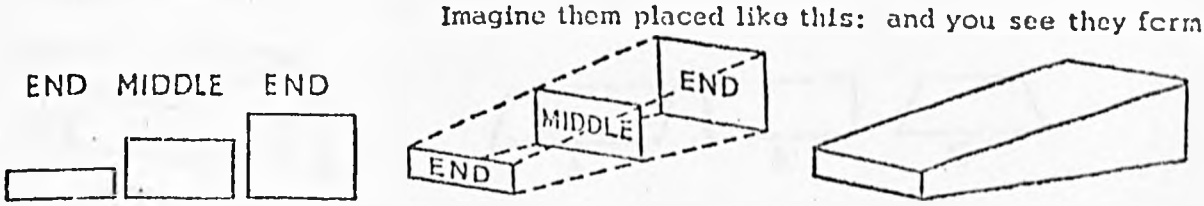
IDENTICAL PICTURES TEST (Identical pictures)

The Identical Pictures Test consists of items similar to the one shown below. The subject is to select from the five figures or pictures on the right the one which is identical to the figure at the left of the row. This is a test of perceptual speed.

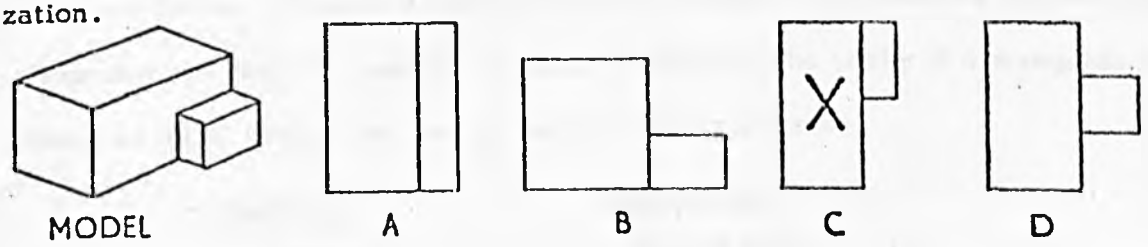
NEWCASTLE SPATIAL TEST (Spatial ability)

The Newcastle Spatial Test, developed by I. McFarlane Smith and J. S. Lawes for the National Foundation for Educational Research in England and Wales, consists of six different sub-tests ranging in difficulty from simple recognition of selections of regular solids to the more complex problems of surface development.

Test 1 consists of ten sets of drawings in which the end views and middle section of a solid object (in the order end, middle, end) are shown. The subject is to determine which one of 12 solid objects on the opposite page fits each set of drawings. (See sample below.) It appears that this test does not require a very strongly developed spatial ability beyond some idea of perspective drawing.

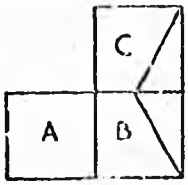


A sample of test 2 is shown below. This test requires the subject to indicate which one of four choices is a view from above of the solid model shown at the left of the row. This test also seems to call for only a modest amount of spatial visualization.

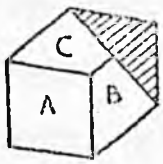


Test 3 consists of items similar to the sample below. In each item the subject is given three sides of a cube in a flat pattern and a drawing of a solid cube, part of which is shaded. The subject is to draw lines on the pattern to indicate where he would cut to remove the parts shown shaded on the solid model. One would

probably use spatial visualization to solve this problem although it seems possible to perform the task by verbal reasoning.

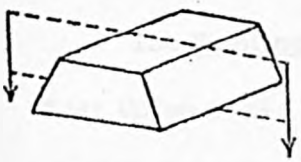


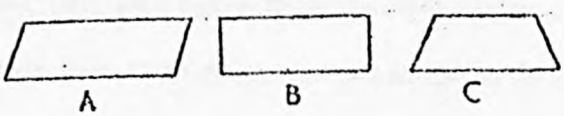
SHAPE



MODEL

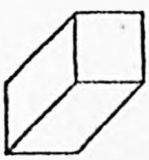
In test 4 each item shows a block of wood. The subject is to imagine a cut made where shown by the dotted lines and to indicate which of the three drawings on the right shows the shape of the cut face. A sample is shown below. It appears that for this task no highly developed ability to visualize 3-dimensional objects is needed.





In each item on test 5 there is a drawing of a solid object, called Shape, and a place to copy it, called Framework. The subject is to put circles around the crosses in the Framework which could be joined to make the Shape. An example is shown below. It appears that it is not necessary to visualize the shape in three dimensions in order to copy it. In fact, the task may be easier if one regards the shape as a flat pattern and merely counts units of distance.

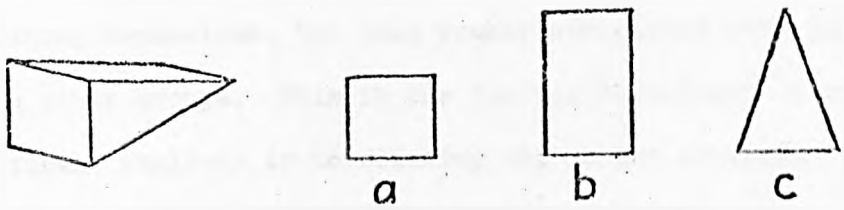
SHAPE E



FRAMEWORK F

1X 2X 3X 4X 5X
1X 2X 3X 4X 5X
1X 2X 3X 4X 5X
1X 2X 3X 4X 5X
1X 2X 3X 4X 5X

Each item in test 6 (see example below) shows a model built from the shapes shown next to it. The subject is required to indicate the number of times each shape was used to make the model. Although one could rely largely on verbal reasoning to solve these problems, visualization would probably allow him to work much faster.



SPELLING ACHIEVEMENT TEST (Spelling)

The Spelling Achievement Test was taken from the Metropolitan Achievement Test (Allen et al, 1946). In this test each word was pronounced by the examiner, used in a sentence, and then pronounced again. The student was then instructed to write the word. The test, consisting of 60 words, was administered to small groups of subjects by trained examiners in accordance with standard procedures. Examples are given below:

garage	I keep my car in a garage.	garage
instructor	One who teaches is an instructor.	instructor
tuberculosis	Tuberculosis is a serious lung disease	tuberculosis

APPENDIX C: THE CONCEPT AND THEORY OF FACTOR ANALYSIS

In psychological and biological studies, many measurements are often made on each individual in the sample. Usually the observed measurements will be correlated. Within the correlation or covariance matrices, the observed variables will tend to fall into groups of variables which are highly correlated among themselves, but only weakly correlated with variables belonging to other groups. This is the "factor structure" of the data. The purpose of factor analysis is to discover the number of common factors underlying the variables and give the covariance between them. The variables with significant specific variation and the proportion of variance accounted for by the specific variation are established. Finally factor analysis shows which factors contribute to variation in which variables allowing inferences about the common causal relationships between the variables.

Factor analysis analyses the original p correlated variables into a small number (m) of uncorrelated variables called common factors, with an uncorrelated residual or specific component remaining for each variable which is also uncorrelated with any of the other variables.

Thus, only a proportion of the variances in the diagonal of the correlation or covariance matrix are assumed to be due to the common factor. This is one advantage of Factor Analysis over Principal Components analysis, which is sometimes used as an approximation to Factor Analysis, in the study of behavioural traits. Principal Components Analysis is superficially similar to Factor analysis, but makes certain assumptions that are untenable in the behavioural context. For example, Principal Components Analysis (PCA) does not allow for some of the variance being specific to certain variables.

The decomposition of the variables into factors does not have a unique solution unless certain restrictions are imposed and the problem of which decomposition to choose is known as the problem of rotation. The clustering of variables into groups or common factors may be considered to occur in

a multidimensional space in which the common factors extracted are the axes. Simple transformations of the factors or, as it is otherwise known, rotation of the axes, will produce an infinite number of solutions from which we have to choose the most appropriate. PCA assumes a unique solution to the problem of rotation such that the orthogonal axes of the principal components solution are the axes of greatest variance. However, the axes of greatest variance are not necessarily the easiest to interpret psychologically. In Factor Analysis, the axes of the common factors are rotated to new orthogonal or oblique axes so that they are clearly interpretable in terms of theoretical ideas about the underlying structure of the traits. This is another advantage of Factor Analysis over Principal Components Analysis as far as determining the underlying structure of psychological variables is concerned.

We must now turn to a more precise formulation of the factor analytic model and a more rigorous discussion of the problems of rotation.

Formulation of the Model

The theory of Factor Analysis has been described by many authors (Joereskog, 1963; Seal, 1964; Harman, 1967; Morrison, 1967; Hope, 1968; Lord and Novick, 1968). We shall follow the notation of Seal in describing the factor analytic model.

Firstly, let p = number of observed variables e.g. the fourteen tests of the Georgia Twin Study

m = the number of underlying factors (i.e. latent variables)

Now, we may define X as a p x 1 row vector of p observed variables, having finite variances, which is N (μ,Σ). The purpose of the factor analysis is to explain the correlational structure of the observed variables in terms of a small number of common factors, the residual variance of each variable being specific. Therefore, let f be a row vector of m common factors with $m < p$. Both X and f are defined in some fixed population of persons.

Now let T be a p x m matrix of weights relating X to f. These weights are known as factor loadings. If we also specify that each columm of T has

at least two non-zero elements then each element of f must be a common factor, since this implies that each such element is common to at least two of the observed variables. The p x 1 vector, μ which is N (0, Δ) and independent of f, contains specific factors corresponding to the elements of X. These are the residuals after the common factors have been extracted. The p x p variance - covariance matrix of μ is called Δ . This is a diagonal matrix with non-negative elements. This implies that Δ really contains specific variances since the elements of Δ are uncorrelated. The matrix TT must be of rank $m < p$.

Now the linear factor analytic model may be written:

$$\underset{(px1)}{X} = \underset{(px1)}{\mu} + \underset{(pxm)(mx1)}{T\ f} + \underset{(px1)}{u} \quad \ldots\ldots\ldots(1)$$

We can decompose the variance-covariance matrix of X, Σ, into two parts.

$$\begin{aligned}\Sigma &= \mathcal{E}\left((x-\mu)(x-\mu)^1\right)\\ &= \mathcal{E}\left((Tf+u)(Tf+u)^1\right)\\ &= \mathcal{E}\left(Tf(Tf)^1 + Tfu^1 + u(Tf)^1 + uu^1\right)\\ &= \mathcal{E}\left(Tff^1T^1\right) + \mathcal{E}(uu^1)\end{aligned}$$

Now since $\mathcal{E}u = 0$ and $\mathcal{E}f = 0$ and u and f are independent,

$$\Sigma = T\,\mathcal{E}\,(ff^1)\,T^1 + \Delta \quad \ldots\ldots\ldots\ldots\ldots(2)$$

This equation is the basis of factor analysis. Many solutions have been proposed which are mostly approximate and not independent of scale change. Kendall (1957) criticises such approximations because the factors and residuals estimated by these methods have unknown distributions and valid tests of significance cannot be made.

We will discuss two commonly used solutions of this equation. However, first there are two points worth making. It has been assumed throughout that the sample of multivariate observations has been drawn from the multivariate Normal distribution and the methods described here should only be applied to data which conforms to this assumption.

If Δ are the specific variances, then

$$T\mathcal{E}\,(ff^1)\,T^1 = \Sigma - \Delta$$

will have diagonal terms which correspond to the common variances. The communalities of the observed variables, X, are defined as the diagonal elements of the matrix, $D^{-1}\,T\,\mathcal{E}\,(ff^1)\,T^{-1}\,D^{-1}$ where $D^{-1} = \text{diag}\,\underset{\sim}{\Sigma}$. The communality of a variable corresponds to the squared multiple correlation coefficient between the variable and the set of common factors.

The Principal Factor Solution

The values of T and Δ must be estimated from the sample of observations, which presents problems since we have two unknowns:

$$\Sigma = T\,(\mathcal{E}\,ff^1)\,T^1 + \Delta$$

The $\underset{\sim}{\Delta}$ obviously cannot be known until the $\underset{\sim}{T}$ has been estimated. Therefore an iterative solution has to be used starting from assumed values of $\underset{\sim}{\Delta}$, or else, likely values of $\underset{\sim}{\Delta}$ can be inserted according to known properties of the data. Therefore, we can write:

$$\hat{\Sigma} - \hat{\Delta} = T\,(\mathcal{E}\,ff^1)\,T^1$$

We can estimate the T from a Principal Components Analysis working with the matrix of $\hat{\Sigma} - \hat{\Delta}$. Any matrix of the form BB^1 is symmetric and all its eigen-values are non-negative. So, if there are negative values among the largest m eigen values of $\hat{\Sigma} - \hat{\Delta}$ then either $\hat{\Delta}$ is an inappropriate estimate or $\hat{\Sigma} - \hat{\Delta}$ cannot be written as $T\,\mathcal{E}\,(ff^1)\,T^1$. Therefore, we write

$$\left|\,(\hat{\Sigma} - \hat{\Delta}) - k\,I\,\right| = 0$$

and proceed to calculate the m largest roots. Next, we calculate $\hat{\hat{\Delta}} = \text{diag}\,(\hat{\Sigma} = T\,(\mathcal{E}f\,f^1)\,T^1)$ and use this as a second estimate of $\hat{\Delta}$. The estimation of T is repeated using new values of $\hat{\Delta}$ until two successive estimates of T change only by a certain specified amount. After completion of the iterations,

$$\text{diag}\quad (T\,\mathcal{E}(ff^1)\,T^1 + \hat{\Delta}) = \text{diag}\,\hat{\Sigma}$$

but,

$$\text{non-diag},\ T\,(\mathcal{E}ff^1)\,T^1 \neq \text{non-diag}\,\hat{\Sigma}$$

Therefore, we can test the "correctness" of m by asking whether the last

(p-m) roots of the final iteration are sufficiently close to zero for the differences between $\hat{\Sigma}$ and $\hat{T}\hat{T}' + \Delta$ to be explained by sampling variation.

The use of Principal Factoring to obtain estimates of T may be criticised because different results are produced when X is re-scaled.

Canonical Factor Analysis

A solution which does not vary according to the scale of X is given by Canonical Factoring. This solves the equation:

$$| \hat{\Delta}^{-\frac{1}{2}} (\hat{\Sigma} - \hat{\Delta}) \hat{\Delta}^{-\frac{1}{2}} - vI | = 0$$

The solution of this equation gives the maximum likelihood solution of Lawley (1940, 1942). An iterative solution was proposed by Maxwell (1961, 1964) using the principal components of the data as the first approximation. The Principal Factoring method has been used in all Factor Analysis carried out in this study.

Rotation of the Factors

The problem of rotation has been briefly described in the introduction to this Appendix. We have pointed out that the decomposition of the p variables into m factors does not have a unique solution. Let T be an orthogonal non-singular matrix of order m, then:

$$\underset{\sim}{T}'\underset{\sim}{T} = \underset{\sim}{I} = \underset{\sim}{T}\underset{\sim}{T}'$$

and

$$(\underset{\sim}{\Psi}\,\underset{\sim}{T})\,(\underset{\sim}{\Psi}\,\underset{\sim}{T})^{1} = \Psi TT'\Psi$$

$$= \Psi\Psi'$$

Therefore, $\Sigma = \Psi\Psi' + \Delta$, irrespective of T. Therefore if a factor analytic solution exists for some value of m, it follows that there are an infinite number of solutions. Going from one solution to another is known as rotation.

The number of possible solutions may be reduced by introducing certain restraints. One such restraint commonly applied is to require that the Ψ common factors are orthogonal and not oblique. This reduces the number of solutions considerably, but does not solve the problem since this restraint will not produce a unique solution. Also, it is often not a very useful

restraint since oblique factors are often found to make more sense, psychologically, than orthogonal factors. Thurstone (1945) has produced a number of criteria which must be satisfied in order to produce "simple structure" of the pxm Ψ matrix of rotated factor loadings. These criteria ensure that variables fall into mutually exclusive groups, having high loadings on one factor, possibly intermediate loadings on a few other factors and virtually no loading on the remaining factors. Thus, Thurstones criteria allow for either orthogonal, or oblique solution of the rotation problem.

Originally, an approximation to simple structure was sought graphically by pair-wise rotation of the factor axes. Morrison (1967) illustrates graphical rotation by means of several examples. When the angle of rotation has been determined, new loadings on the rotated factors may be obtained approximately as the projections of the variables on the rotated axes, or may be determined directlyby post-multiplying the factor loadings on the old axes by the *transformation matrix:*

$$T = \begin{pmatrix} \cos & \sin \\ -\sin & \cos \end{pmatrix}$$

However, graphical rotation is very tedious and time-consuming, especially as the number of factors and variables increases. Therefore, later workers have proposed several analytical solutions for orthogonal or oblique rotation, using various criteria in an attempt to approximate simple structure. Ferguson (1954) and Nehaus and Wrigley (1954) independently produced solutions for orthogonal rotation based on what is now known as the quartimax criterion for rotation. Quartimax rotation attempts to maximise the fourth powers of the factor loadings. Carroll (1953) also used the quartimax criterion for achieving simple structure, but his formulation permitted oblique as well as orthogonal solutions.

Kaiser (1958) proposed a different criterion for producing simple

structure, known as the varimax criterion, which attempts to maximise the squared loadings in each column of the factor matrix. Harman (1967) has made a comparison of the loadings obtained through quartimax and varimax rotation, and shown that the varimax criterion produces solutions which more nearly approximate simple structure than the quartimax criterion. Hendrickson and White (1964) have proposed an alternative method of oblique rotation to simple structure known as the Promax method.

APPENDIX D: MULTIPLE REGRESSION ANALYSIS

The most general form of multiple regression analysis will be discussed here, where q error free measurements are used to explain p interconnected variates. A full discussion of multiple regression analysis may be found in Seal (1968), and we follow Seal in writing the full model as:

$$
\begin{aligned}
x_{ij} &= \beta_{10} z_{0j} + \beta_{11} z_{ij} + \beta_{12} z_{2j} + \beta_{13} z_{3j} + \cdots \beta_{1q} z_{qj} + e_{ij} \\
x_{2j} &= \beta_{20} z_{0j} + \beta_{21} z_{ij} + \beta_{22} z_{2j} + \beta_{23} z_{3j} + \cdots \beta_{2q} z_{qj} + e_{2j} \\
\vdots \quad & \qquad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\
x_{pj} &= \beta_{p0} z_{0j} + \beta_{p1} z_{0j} + \beta_{p2} z_{2j} + \beta_{p3} z_{3j} + \cdots \beta_{pq} z_{qj} + e_{pj}
\end{aligned}
$$

where x_p are the p-variates.

β_{pj} are the pj regression coefficients for the p variates and the q measurements.

j is the number of units on which the measurements are made - usually the number of individuals.

e_{pj} are the error terms.

N.B. The first term on the right hand sides (i.e. z_{oj} is the mean value and the following terms up to z_{qj} represent the explanatory variables. The p component vectors of x measurements are:

$$
x_1 = \begin{bmatrix} x_{11} \\ x_{12} \\ x_{13} \\ \cdot \\ \cdot \\ \cdot \\ x_{1j} \end{bmatrix} \qquad x_2 = \begin{bmatrix} x_{21} \\ x_{22} \\ x_{23} \\ \cdot \\ \cdot \\ \cdot \\ x_{2j} \end{bmatrix} \quad \cdots \quad x_p = \begin{bmatrix} x_{p1} \\ x_{p2} \\ x_{p3} \\ \cdot \\ \cdot \\ \cdot \\ x_{pj} \end{bmatrix}
$$

Let $X = [x_1\ x_2 \ldots\ldots x_p]$

The same notation is used for e_1, e_2 and E

Now $Z = \begin{bmatrix} z_{01} & z_{02} & \cdots\cdots z_{0j} \\ z_{11} & z_{12} & \cdots\cdots z_{1j} \\ z_{21} & z_{22} & \cdots\cdots z_{2j} \\ \vdots & \vdots & \vdots \\ z_{q1} & z_{q2} & \cdots\cdots z_{qj} \end{bmatrix}$

and $\beta_1 = \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1q} \end{bmatrix}$ $\beta_2 = \begin{bmatrix} \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \vdots \\ \beta_{2q} \end{bmatrix}$ $\beta_p = \begin{bmatrix} \beta_{p0} \\ \beta_{p1} \\ \beta_{p2} \\ \vdots \\ \beta_{pq} \end{bmatrix}$

Therefore, following the previous notation

$$B = [\beta_1 \quad \beta_2 \cdots\cdots \beta_p]$$

The full model may now be re-written as follows:

$$\underset{(jxp)}{X} = \underset{(jxq)}{Z'} \quad \underset{(qxp)}{B} + \underset{(jxp)}{E}$$

Now premultiplying by Z and replacing B by $\hat{B}$ its least squares estimate (NB $\Sigma(E) = 0$ and $\Sigma(E'E) = \Sigma$, where 0 is a jxp matrix of zeros and

$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \cdots\cdots \sigma_{1p} \\ \sigma_{21} & \sigma_{22} \cdots\cdots \\ \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} \cdots\cdots \sigma_{pp} \end{bmatrix}$, i.e

the variance - covariance matrix), we now obtain:

$$Z \quad X = Z \quad Z' \quad \hat{B}$$

from whence

$$\hat{B} = (ZZ')^{-1} \; Z \; X$$

This full model is designated Ω and assumes that all q measurements are affecting the p-variates. The next stage is to eliminate s rows of the B_{Ω} and write the reduced models designated by W as:

$$\underset{(jxp)}{X} = \underset{(jx(q-s))}{Z'} \quad \underset{((q-s)x\ p)}{B} + \underset{(jxp)}{E}$$

The s rows which are eliminated depend on the design of the experiment and the effects we are interested in. When the W model has been written we proceed to test the validity of the reduced model by comparing the two sets of residuals from Ω and W respectively. We calculated $(N-q)\hat{\Sigma}_{\Omega}$ and $(N-S)\hat{\Sigma}_{W}$ where

$$f\hat{\Sigma} = X'X - \hat{B}'ZZ'\hat{B} \quad \text{(See Seal)}$$

and B and Z refer to either Ω or W provided f is adjusted appropriately. The ratio of the determinants is written $U_{p,q-s,f}$ and the variate $-[f-\frac{1}{2}(p-q\ s+1)]\ \ln U_{p,q-s,f}$ is distributed approximately as chisquare with p(q-s) degrees of freedom. If the chisquare is significant, then the significance of the p(q-s) values of B omitted from the model are considered to be significant.

This technique is often useful in psychological studies when different psychological variables can be partly explained in terms of age, socio-economic status and other similar causes. It enables a correction to be made to the data for the predictor variables so that variation in the variable due to other causes can be estimated without being biased by the effects of the criterion variables.

APPENDIX E: MAXIMUM-LIKELIHOOD ESTIMATION AND WEIGHTED LEAST SQUARES.

MAXIMUM-LIKELIHOOD ESTIMATION

A problem exists when we wish to estimate population parameters from a sample and several estimators exist. The criteria used to determine the "best" estimator are based on the distributions of estimates produced by the estimators. Four properties of estimators are important in discriminating between them and in choosing the "best" estimator.

CONSISTENCY

Consistency is a limiting property, affecting the behaviour of estimators (t's) as sample size (n) approaches infinity. A consistent estimator is one where the probability that estimates (θ's) are close to the true values of the parameters (θ's) approaches unity, as the sample size tends to infinity.

Lack of Bias

An unbiased estimator is one where:

$$\mathcal{E}(t) = \theta$$

for all values of n and θ. Thus, a consistent estimate may be biased, but consistent estimators must tend to be unbiased as the sample size approaches infinity. It is possible that there may be no unbiased estimator for a parameter.

Efficiency

If we choose consistent, unbiased estimators, then estimators with a smaller variance will deviate less, on average, from the true parameter value, than estimators with a larger variance. This is known as the criterion of minimum variance and an estimator with minimum variance as the sample size tends to infinity, is known as an efficient estimator.

Sufficiency

Fisher (1921,1925) defined the criterion of the sufficiency of estimators. A sufficient estimator is one which uses all the information in the sample about the parameter being estimated.

In order to find an estimator that satisfied these criteria, the method of Maximum-likelihood is used. Kendall and Stuart (1961) outline the principles and show how the best estimator of an effect may be found by the method of maximum-likelihood. Fisher has shown that the maximum-likelihood estimator satisfies most of the criteria specified above. Maximum-Likelihood estimators are consistent, although they are not necessarily unbiased. They are sufficient, where a sufficient estimator exists, and have minimum variance in larges amples (i.e. they are efficient) and sometimes in small samples also. The Maximum-Likelihood estimator is asymptotically normally distributed (i.e. its distribution approaches normality as the sample size approaches infinity) and shows the property of functional invariance (i.e. if $\hat{\theta}$ is the maximum-likelihood estimate of θ, then $f(\hat{\theta})$ is the maximum-likelihood estimate of $f(\theta)$).

WEIGHTED LEAST SQUARES

When we wish to explain a set of statistics in terms of a linear combination of the parameters of a model, then the method of least squares provides maximum-likelihood estimates of the parameters, given certain assumptions.

These assumptions are:

1. That the observations are independent
2. That the observations are normally distributed

When the statistics are each based on the same number of observations, then least squares is used. Where the sample size differs with each statistic, the statistics are not known with equal precision and estimates of the parameters obtained by least squares are not independent. However, this problem may be overcome and maximum-likelihood estimates of the parameters obtained by appropriately weighting the statistics. In the method of weighted least squares, the statistics are weighted by the inverse of their variances. For normally distributed estimates, the inverse of the variance is equivalent to the amount of information about that estimate.

We may now formulate the linear equation for which we wish to find the least squares estimator:

$$\underset{\sim}{X} = \underset{\sim}{A} \quad \underset{\sim}{W} \quad \underset{\sim}{\theta} \quad + \quad \underset{\sim}{e}$$

where $\underset{\sim}{X}$ is an n x 1 vector of statistics

$\underset{\sim}{A}$ is an n x k matrix of the coefficients of the model

$\underset{\sim}{W}$ is an n x n diagonal matrix of weights

θ is a k x 1 vector of the parameters of the model

and $\underset{\sim}{e}$ is an n x 1 vector of error terms

The expected value of $\underset{\sim}{e}$, $\mathcal{E}(\underset{\sim}{e}) = 0$, and the variance-covariance matrix of $\underset{\sim}{e}$ is given by $V(\underset{\sim}{e}) = \mathcal{E}(\underset{\sim}{e}\,\underset{\sim}{e}') = \sigma^2 \underset{\sim}{I}$ where $\underset{\sim}{I}$ is the n x n identity matrix. Thus, we are assuming that the e's are uncorrelated, have zero means and the same variance.

To obtain the least squares estimator of θ, we must minimise the sum of squares of $\underset{\sim}{X} - \underset{\sim}{A}\,\underset{\sim}{W}\,\underset{\sim}{\theta}$ for variation in θ, which is given by:

$$\text{Sum of Squares} = S = (\underset{\sim}{X} - \underset{\sim}{A}\,\underset{\sim}{W}\,\underset{\sim}{\theta})'(\underset{\sim}{X} - \underset{\sim}{A}\,\underset{\sim}{W}\,\theta)$$

In order to minimise S, we must let $\frac{ds}{d\theta} = 0$.

Differentiating, we find:

$$2\underset{\sim}{A}'\underset{\sim}{W}\,(\underset{\sim}{X} - \underset{\sim}{A}\,\underset{\sim}{W}\,\underset{\sim}{\theta}) = 0$$

which gives the least squares estimator as:

$$\underset{\sim}{\hat{\theta}} = (\underset{\sim}{A}'\underset{\sim}{W}\,\underset{\sim}{A})^{-1}(\underset{\sim}{A}'\underset{\sim}{W}\,\underset{\sim}{X}) \quad \ldots\ldots\ldots (1)$$

The matrix (A'W A) is the weighted information matrix. This matrix comprises the weighted sums of squares and products of the elements of the column vectors of A, the model matrix. The inverse of the information matrix yields the variance - covariance matrix of the estimates. The parameter estimates are obtained by evaluating equation (1). The variances of these estimates are given by the diagonal terms of the variance-covariance matrix. The estimates may be tested for significant departures from zero by use of the c test. The one-tailed test is used when we expect an estimate to take a particular sign. In the absence of any expectation for the sign of a parameter estimate,

the two-tailed test is used.

If parameter estimates turn out to be non-significant, the expectations for these parameters may be dropped from the model and a simpler model may be fitted. We now have the problem of deciding which of our models are adequate to explain variation in our observations. The "goodness of fit" of a given model may be found since the sum of squares of the residuals (i.e. the $\underset{\sim}{e}$'s), $(n-k)s^2/\sigma^2$, is a chisquare with n-k degrees of freedom. If this chisquare is significant, the model we have fitted is not adequate to explain variation in the trait we are interested in and further models must be tested.

APPENDIX G: TABLES OF BIASES AND POWER

This appendix contains twelve tables, providing information about fourteen types of population, each with different genetical and environmental causes of variation. The population parameters have been specified in the text of part C. The meaning of the symbols used for the parameters is briefly summarised in Key A and their contribution to the total variation is also given. All the keys are placed immediately before the Tables.

The genetical and environmental effects contributing to variation in each of the fourteen population types are summarised in Key B. Key B also indicates the number of the Table in which imformation about a particular population can be found.

The fourteen populations have been classified into five groups, whose main interests are additive effects, dominance, genotype-environment interaction, cultural transmission and sibling effects. This classification forms the basis of the discussion in the text.

A number of models have been fitted to each population, using the methods described in Section 3 of Part A and Appendix E, in order to determine:

1. The contribution of "true" population parameters to parameters estimated using the "wrong" model i.e. the biases which may occur in model fitting.
2. How heritabilities may be biased when the wrong model is fitted.
3. The number of individuals required for a given estimated parameter to be significant at the 5% level in 95% of cases.

4. The number of individuals required for rejection of the wrong model at the 5% level to occur in 95% of cases.

The models which were used and appear in the Tables are summarised in Key C. The information obtained is presented in two types of table, labelled A and B. Additional numbering is used when more than one page is required for any table e.g. Table 1B.2 refers to the second page of Table 1B.

The "A" Tables

These Tables contain the biases introduced into parameter estimates when a wrong model is fitted to a particular population type. The contribution of the "true" population parameters to the parameters estimated using the wrong model is tabulated. The "true" parameters form the column headings. The models fitted to the population are numbered downwards. The estimated parameters of the wrong model can be read in terms of the "true" parameters e.g. when the E_1E_2 model is wrongly fitted to a population in which the true causes of variation are E_1 and D_R, it can be seen from Table 1A that E_1 would be estimated as $E_1 + \frac{1}{2} D_R$ and E_2 as $\frac{1}{2} D_R$.

The following additional points must be noted.

1. No decimal places are given when the contribution of the "true" parameter to the estimated parameter is an integer.
2. In other cases, two decimal places only are given.
3. A value of 0.00 indicates that the contribution of the "true" parameter to the estimated parameter is less than 0.005.
4. A " - " indicates no contribution of the "true" parameter to the estimated parameter.

The "B" Tables

1. The columns headed "True Parameter Values" give the variance produced by the true population parameters, given particular levels of heritability, assortative mating, dominance, genotype-environment interaction, genotype-environment covariance produced by cultural transmission and genotype-environment covariance produced by sibling effects. For any population type there are a series of different populations in which the variation produced by the true parameters differs. These are labelled a, b, c, etc. in Tables 1-10. In Tables 11 and 12, the different populations are characterised by the two parameters h^2_N and ρ, where h^2_N is the narrow heritability and ρ is the proportion of loci which contribute both to D_R and D_R'' i.e. the proportion of loci having both direct effects on the phenotype and indirect effects on the phenotype of siblings (see text for conditions under which this is true). How the different populations were specified is explained in the text of Part C.

2. The models fitted to the populations are numbered downwards, and their parameters can be determined as the parameters under which there are entries in the columns headed 'Sample Sizes Needed to Detect:"

3. The entries in these columns give the number of individuals required for estimated parameters to be significant at the 5% level in 95% of cases.

4. The number of individuals required for rejection of a wrong model at the 5% level in 95% of cases is given in the column headed "R".

5. Estimates of narrow (h^2_N) and broad (h^2_B) heritability are also given, where h^2_N is defined as:

$$h^2_N = \frac{\frac{1}{2} D_R + \frac{1}{2}M}{V_T}$$

and h^2_B is defined, somewhat unconventionally as:

$$h^2_B = \frac{\frac{1}{2} D_R + \frac{1}{2}M + \frac{1}{4}H_R + \frac{1}{2} I + C + \frac{1}{2} D_R'' + D_R'}{V_T}$$

and V_T is the total variance.

The following additional points must be noted.

a. Models marked by an asterix are the correct models for the population to which they are being fitted.

b. Where the sample size needed to detect an effect is followed by a "X", the estimated value of the effect is negative.

c. A value given as "∞", indicates that the number was greater than 99,999.

KEY A: SYMBOLS USED FOR GENETICAL AND ENVIRONMENTAL PARAMETERS

SYMBOL	COMTRIBUTION TO TOTAL VARIATION	MEANING OF SYMBOL
E_1	1	Specific environmental variation
E_2	1	Common environmental variation
D_R	$\frac{1}{2}$	Additive genetical variation
M	$\frac{1}{2}$	Additive genetical variation produced by assortative mating
H_R	$\frac{1}{4}$	Dominance variation
I	$\frac{1}{2}$	Systematic genotype-environment interaction
C	1	Variation produced by cultural transmission
D_R''	$\frac{1}{2}$	Variation produced by the contribution of the additive genetical deviations of siblings to the environment of their co-siblings
D_R'	1	Covariance between genotype and environment produced by the contribution of the phenotypes of sibs to the environments of their co-sibs.

KEY B: GENETICAL AND ENVIRONMENTAL FACTORS CONTRIBUTING TO VARIATION IN FOURTEEN POPULATIONS

POPULATION	TABLE	POPULATION PARAMETERS	
1	1	E_1 D_R	Additive effects
2	1	E_1 E_2 D_R	
3	1	E_1 D_R M	
4	2	E_1 E_2 D_R M	
5	3	E_1 D_R H_R	Dominance
6	4	E_1 E_2 D_R H_R	
7	5	E_1 D_R H_R M	
8	6	E_1 E_2 D_R H_R M	
9	7	E_1 E_2 D_R I	Genotype-environment interaction
10	8	E_1 E_2 D_R I M	
11	9	E_1 E_2 D_R H_R I	
12	10	E_1 E_2 D_R C	Cultural transmission
13	11	E_1 D_R D_R'' D_R'	Sibling effects
14	12	E_1 E_2 D_R D_R'' D_R'	

KEY C: MODELS FITTED TO THE POPULATIONS

$E_1\ E_2$

$E_1\ D_R$

$E_1\ E_2\ D_R$

$E_1\ D_R\ M$

$E_1\ D_R\ H_R$

$E_1\ E_2\ D_R\ M$

$E_1\ E_2\ D_R\ H_R$

$E_1\ D_R\ H_R\ M$

$E_1\ E_2\ D_R\ I$

$E_1\ E_2\ D_R\ C$

$E_1\ D_R\ D_R''\ D_R'$

$E_1\ E_2\ D_R\ H_R\ M$

$E_1\ E_2\ D_R\ H_R\ I$

$E_1\ E_2\ D_R\ D_R''\ D_R'$

$E_1\ D_R\ D_R''\ D_R'\ M$

TABLE 1A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS		
		E_1	D_R	
1	E_1	.1	0.25	
	E_2	-	0.25	
		E_1	E_2	D_R
1	E_1	1	-	0.25
	E_2	-	1	0.25
2	E_1	1	0.46	-
	D_R	-	1.05	1
3	E_1	1	0.5	-
	D_R	-	3	1
	H_R	-	-	-4
4	E_1	1	0.51	-
	D_R	-	-0.50	1
	M	-	1.34	-
		E_1	D_R	M
1	E_1	1	0.25	0.15
	E_2	-	0.25	0.39
2	E_1	1	-	-0.04
	D_R	-	1	1.16
3	E_1	1	-	-0.09
	E_2	-	-	0.12
	D_R	-	1	1.03
4	E_1	1	-	0.01
	D_R	-	1	3.18
	H_R	-	-	-4.17

TABLE 1B.1

MODEL	TRUE PARAMETER VALUES			SAMPLE SIZES NEEDED TO DETECT:					R	h_N^2	h_B^2
	E_1	E_2	D_R	E_1	E_2	D_R	H_R	M			
1a	112.50	-	225.00	84	728	-	-	-	562	-	-
b	22.50	-	405.00	90	205	-	-	-	147	-	-
2a	112.50	-	225.00	117	-	191	-	-	-	0.50	-
b	22.50	-	405.00	141	-	41	-	-	-	0.90	-
1a*	56.25	56.25	225.00	91	160	-	-	-	440	-	-
b*	11.25	11.25	405.00	91	160	-	-	-	137	-	-
2a	56.25	56.25	225.00	127	-	105	-	-	1085	0.63	-
b	11.25	11.25	405.00	142	-	38	-	-	2215	0.92	-
3a*	56.25	56.25	225.00	221	633	207	-	-	-	0.50	-
b*	11.25	11.25	405.00	282	1361	42	-	-	-	0.90	-
4a	56.25	56.25	225.00	139	-	1114	$13444^{x}_{\infty x}$	-	1010	0.88	0.63
b	11.25	11.25	405.00	144	-	749		-	2151	0.98	0.93
5a	56.25	56.25	225.00	132	-	1040	-	4544	1237	0.45	0.62
b	11.25	11.25	405.00	142	-	157	-	89594	2171	0.89	0.92

TABLE 1B.2

MODEL	TRUE PARAMETER VALUES			SAMPLE SIZES NEEDED TO DETECT:					R	h_N^2	h_B^2
	E_1	D_R	M	E_1	E_2	D_R	H_R	M			
1a	112.5	225.0	75.0	86	424	-	-	-	401	-	-
b	22.5	405.0	331.4	92	123	-	-	-	108	-	-
2a	112.5	225.0	75.0	122	-	126	-	-	12066	0.59	-
b	22.5	405.0	331.4	142	-	36	-	-	948	0.94	-
3a	112.5	225.0	75.0	176	39354	174	-	-	14315	0.57	-
b	22.5	405.0	331.4	284	67914	37	-	-	919	0.94	-
4a	112.5	225.0	75.0	136	-	1160	10244^x	-	28214	0.87	0.58
b	22.5	405.0	331.4	144	-	142	400^x	-	3549	1.84	0.95
5a*	112.5	225.0	75.0	129	-	1222	-	6771	-	0.57	-
b*	22.5	405.0	331.4	144	-	207	-	377	-	0.94	-

TABLE 2A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	M
1	E_1	1	-	0.25	0.16
	E_2	-	1	0.25	0.38
2	E_1	1	0.47	-	-0.03
	D_R	-	1.02	1	1.12
3	E_1	1	-	-	-0.04
	E_2	-	1	-	0.09
	D_R	-	-	1	0.94
4	E_1	1	0.50	-	0.00
	D_R	-	3	1	3.24
	H_R	-	-4	-	-4.27
5	E_1	1	0.51	-	-
	D_R	-	-0.37	1	-
	M	-	1.22	-	1
6	E_1	1	0.52	-	-0.03
	D_R	-	0.84	1	1.01
	D_R''	-	-0.49	-	0.08
	D_R'	-	0.86	-	0.17
7	E_1	1	-	-	-0.01
	E_2	-	1	-	0.05
	D_R	-	-	1	3.02
	H_R	-	-	-	-4.02
8	E_1	1	-	-	-0.01
	E_2	-	1	-	0.26
	D_R	-	-	1	1.16
	I	-	-	-	-0.58
9	E_1	1	0.49	-	-
	D_R	-	-4.26	1	-
	H_R	-	5.57	-	-
	M	-	2.25	-	1

TABLE 2B.1

MODEL	TRUE PARAMETER VALUES				SAMPLE SIZES NEEDED TO DETECT:						R	h_N^2	h_B^2
	E_1	E_2	D_R	M	E_1	E_2	D_R	H_R	D_R''	D_R'			
1a	56.25	56.25	225.00	75.00	91	140	-	-	-	-	313	-	-
b	11.25	11.25	405.00	331.36	92	110	-	-	-	-	98	-	-
2a	56.25	56.25	225.00	75.00	130	-	83	-	-	-	904	0.70	-
b	11.25	11.25	405.00	331.36	142	-	35	-	-	-	611	0.96	-
3a	56.25	56.25	225.00	75.00	238	604	154	-	-	-	11680	0.56	-
b	11.25	11.25	405.00	331.36	286	1221	37	-	-	-	890	0.94	-
4a	56.25	56.25	225.00	75.00	140	-	513	2582[x]	-	-	1036	1.19	0.68
b	11.25	11.25	405.00	331.36	144	-	133	358[x]	-	-	1281	1.88	0.96
5a	56.25	56.25	225.00	75.00	132	-	170	-	70304[x]	982	1545	0.68	-
b	11.25	11.25	405.00	331.36	144	-	79	-	62248	1166	753	0.95	-

TABLE 2B.2

MODEL	TRUE PARAMETER VALUES				SAMPLE SIZES NEEDED TO DETECT:						R	h_N^2	h_B^2
	E_1	E_2	D_R	M	E_1	E_2	D_R	H_R	I	M			
6a	56.25	56.25	225.00	75.00	135	-	1099	-	-	1533	1334	0.67	-
b	11.25	11.25	405.00	331.36	144	-	188	-	-	339	2195	0.96	-
7a	56.25	56.25	225.00	75.00	256	665	1020	7722[x]	-	-	44181	0.85	0.57
b	11.25	11.25	405.00	331.36	285	1289	140	382[x]	-	-	3579	1.84	0.94
8a	56.25	56.25	225.00	75.00	273	1044	190	-	21113[x]	-	15962	0.59	0.51
b	11.25	11.25	405.00	331.36	286	1078	39	-	1357[x]	-	1310	0.94	0.70
9a*	56.25	56.25	225.00	75.00	256	802	822	-	-	6133	-	0.57	-
b*	11.25	11.25	405.00	331.36	285	1393	185	-	-	363	-	0.94	-
10a	56.25	56.25	225.00	75.00	140	-	∞	34466	-	3815[x]	1300	0.37	0.67
b	11.25	11.25	405.00	331.36	144	-	39430	∞	-	3097	2107	0.91	0.96

TABLE 3A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS		
		E_1	D_R	H_R
1	E_1	1	0.25	0.15
	E_2	-	0.25	0.10
2	E_1	1	-	0.02
	D_R	-	1	0.47
3	E_1	1	-	0.02
	E_2	-	-	-0.01
	D_R	-	1	0.49
4	E_1	1	-	0.01
	D_R	-	1	0.63
	M	-	-	-0.13

TABLE 3B

MODEL	TRUE PARAMETER VALUES			SAMPLE SIZES NEEDED TO DETECT:					R	h_N^2	h_B^2
	E_1	D_R	H_R	E_1	E_2	D_R	H_R	M			
1a	112.5	150.0	150.0	84	820	-	-	-	587	-	-
b	22.5	270.0	270.0	89	232	-	-	-	156	-	-
c	112.5	200.0	50.0	84	756	-	-	-	567	-	-
d	22.5	360.0	90.0	90	213	-	-	-	152	-	-
2a	112.5	150.0	150.0	116	-	199	-	-	71089	0.49	-
b	22.5	270.0	270.0	141	-	41	-	-	13405	0.90	-
c	112.5	200.0	50.0	117	-	193	-	-	∞	0.50	-
d	22.5	360.0	90.0	141	-	41	-	-	∞	0.90	-
3a	112.5	150.0	150.0	153	∞ x	260	-	-	72038	0.49	-
b	22.5	270.0	270.0	270	∞ x	42	-	-	13048	0.90	-
c	112.5	200.0	50.0	156	∞ x	255	-	-	∞	0.50	-
d	22.5	360.0	90.0	270	∞ x	42	-	-	∞	0.90	-
4a*	112.5	150.0	150.0	134	-	9505	40853	-	-	0.33	0.50
b*	22.5	270.0	270.0	144	-	2311	8464	-	-	0.60	0.90
c*	112.5	200.0	50.0	134	-	5208	∞	-	-	0.44	0.50
d*	22.5	360.0	90.0	144	-	1206	68433	-	-	0.80	0.90
5a	112.5	150.0	150.0	123	-	863	-	78126 x	∞	0.50	-
b	22.5	270.0	270.0	142	-	151	-	12976 x	38499	0.90	-
c	112.5	200.0	50.0	123	-	951	-	∞ x	∞	0.50	-
d	22.5	360.0	90.0	142	-	159	-	∞ x	∞	0.90	-

TABLE 4A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	H_R
1	E_1	1	-	0.25	0.15
	E_2	-	1	0.25	0.10
2	E_1	1	0.46	-	0.01
	D_R	-	1.05	1	0.49
3	E_1	1	-	-	0.01
	E_2	-	1	-	-0.01
	D_R	-	-	1	0.51
4	E_1	1	0.50	-	-
	D_R	-	3	1	-
	H_R	-	-4	-	1
5	E_1	1	0.53	-	0.01
	D_R	-	0.80	1	0.49
	D_R''	-	0.46	-	0.02
	D_R'	-	0.85	-	-0.02
6	E_1	1	0.51	-	0.00
	D_R	-	0.55	1	0.67
	M	-	1.38	-	-0.16
7	E_1	1	-	-	0.01
	E_2	-	1	-	-0.04
	D_R	-	-	1	0.47
	I	-	-	-	0.09
8	E_1	1	-	-	0.00
	E_2	-	1	-	0.01
	D_R	-	-	1	0.68
	M	-	-	-	-0.18
9	E_1	1	0.49	-	-
	D_R	-	-4.05	1	-
	H_R	-	5.21	-	1
	M	-	2.23	-	-

TABLE 4B.1

MODEL	TRUE PARAMETER VALUES				SAMPLE S	
	E_1	E_2	D_R	H_R	E_1	E_2
1a	56.25	56.25	150.00	150.00	91	171
b	56.25	56.25	200.00	50.00	91	164
c	11.25	11.25	270.00	270.00	90	181
d	11.25	11.25	360.00	90.00	91	168
2a	56.25	56.25	150.00	150.00	127	-
b	56.25	56.25	200.00	50.00	127	-
c	11.25	11.25	270.00	270.00	142	-
d	11.25	11.25	360.00	90.00	142	-
3a	56.25	56.25	150.00	150.00	218	677
b	56.25	56.25	200.00	50.00	221	408
c	11.25	11.25	270.00	270.00	282	1415
d	11.25	11.25	360.00	90.00	282	1379
4a	56.25	56.25	150.00	150.00	139	-
b	56.25	56.25	200.00	50.00	139	-
c	11.25	11.25	270.00	270.00	144	-
d	11.25	11.25	360.00	90.00	144	-
5a	56.25	56.25	150.00	150.00	128	-
b	56.25	56.25	200.00	50.00	128	-
c	11.25	11.25	270.00	270.00	142	-
d	11.25	11.25	360.00	90.00	142	-

SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
D_R	H_R	D_R''	D_R'			
-	-	-	-	464	-	-
-	-	-	-	450	-	-
-	-	-	-	147	-	-
-	-	-	-	137	-	-
108	-	-	-	1154	0.63	-
107	-	-	-	1110	0.63	-
38	-	-	-	2024	0.93	-
38	-	-	-	2229	0.93	-
208	-	-	-	49593	0.50	-
208	-	-	-	∞	0.50	-
42	-	-	-	12124	0.90	-
42	-	-	-	∞	0.90	-
1800	∞ x	-	-	1120	0.71	0.63
1296	22779 x	-	-	1105	0.82	0.63
1760	11524	-	-	2161	0.68	0.93
969	∞	-	-	2156	0.88	0.93
217	-	41965 x	1363	1868	0.61	-
213	-	36568 x	1217	1892	0.61	-
85	-	∞ x	∞	1877	0.93	-
85	-	∞ x	62658	2112	0.93	-

TABLE 4B.2

MODEL	TRUE PARAMETER VALUES				SAMPLE SIZES NEEDED TO DETECT:						R	h_N^2	h_B^2
	E_1	E_2	D_R	H_R	E_1	E_2	D_R	H_R	I	M			
6a	56.25	56.25	150.00	150.00	132	-	868	-	-	9112	1222	0.61	-
b	56.25	56.25	200.00	50.00	132	-	975	-	-	5629	1237	0.61	-
c	11.25	11.25	270.00	270.00	142	-	144	-	-	29198[x]	2034	0.93	-
d	11.25	11.25	360.00	90.00	142	-	170	-	-	∞	2151	0.93	-
7a*	56.25	56.25	150.00	150.00	249	668	8184	30420	-	-	-	0.33	0.50
b*	56.25	56.25	200.00	50.00	249	661	4427	∞	-	-	-	0.44	0.50
c*	11.25	11.25	270.00	270.00	284	1369	2232	7968	-	-	-	0.60	0.90
d*	11.25	11.25	360.00	90.00	284	1364	1162	64018	-	-	-	0.80	0.90
8a	56.25	56.25	150.00	150.00	268	1829	326	-	∞	-	56711	0.49	0.52
b	56.25	56.25	200.00	50.00	269	1572	315	-	∞	-	∞	0.50	0.51
c	11.25	11.25	270.00	270.00	286	∞	43	-	42457	-	14324	0.90	0.95
d	11.25	11.25	360.00	90.00	286	84269	43	-	∞	-	∞	0.90	0.92
9a	56.25	56.25	150.00	150.00	250	713	625	-	-	42184[x]	∞	0.50	-
b	56.25	56.25	200.00	50.00	250	722	700	-	-	∞[x]	∞	0.50	-
c	11.25	11.25	270.00	270.00	285	1340	141	-	-	11588[x]	38622	0.90	-
d	11.25	11.25	360.00	90.00	285	1363	150	-	-	89691[x]	∞	0.90	-
10a	56.25	56.25	150.00	150.00	139	-	∞[x]	11798	-	6146	1237	0.11	0.62
b	56.25	56.25	200.00	50.00	139	-	∞	19049	-	6140	1217	0.22	0.62
c	11.25	11.25	270.00	270.00	144	-	19753	16544	-	∞	2097	0.55	0.92
d	11.25	11.25	360.00	90.00	144	-	9170	76414	-	∞	2092	0.75	0.92

TABLE 5A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	D_R	H_R	M
1	E_1	1	0.25	0.15	0.15
	E_2	-	0.25	0.10	0.40
2	E_1	1	-	0.01	-0.05
	D_R	-	1	0.48	1.18
3	E_1	1	-	0.02	-0.11
	E_2	-	-	-0.11	0.13
	D_R	-	1	0.49	1.04
4	E_1	1	-	-	0.01
	D_R	-	1	-	3.11
	H_R	-	-	1	-4.06
5	E_1	1	-	0.01	-0.03
	D_R	-	1	0.48	1.05
	D_R''	-	-	0.02	0.03
	D_R'	-	-	-0.02	0.16
6	E_1	1	-	0.01	-
	D_R	-	1	0.64	-
	M	-	-	-0.14	1
7	E_1	1	-	-	-0.03
	E_2	-	-	-	0.09
	D_R	-	1	-	2.85
	H_R	-	-	1	-3.71
8	E_1	1	-	0.01	-0.02
	E_2	-	-	-0.03	0.30
	D_R	-	1	0.48	1.19
	I	-	-	0.06	-0.67
9	E_1	1	-	0.00	-
	E_2	-	-	0.01	-
	D_R	-	1	0.65	-
	M	-	-	-0.15	1

TABLE 5B.1

MODEL	TRUE PARAMETER VALUES				SAMPLE SIZES		
	E_1	D_R	H_R	M	E_1	E_2	D_R
1a	112.50	150.00	150.00	30.00	84	621	-
b	22.50	270.00	270.00	115.71	91	171	-
c	112.50	200.00	50.00	57.14	85	482	-
d	22.50	360.00	90.00	240.00	91	136	-
2a	112.50	150.00	150.00	30.00	120	-	164
b	22.50	270.00	270.00	115.71	142	-	38
c	112.50	200.00	50.00	57.14	121	-	139
d	22.50	360.00	90.00	240.00	144	-	37
3a	112.50	150.00	150.00	30.00	164	∞	219
b	22.50	270.00	270.00	115.71	276	∞	39
c	112.50	200.00	50.00	57.14	172	72130	189
d	22.50	360.00	90.00	240.00	280	88286	38
4a	112.50	150.00	150.00	30.00	135	-	3934
b	22.50	270.00	270.00	115.71	144	-	537
c	112.50	200.00	50.00	57.14	136	-	1673
d	22.50	360.00	90.00	240.00	144	-	207
5a	112.50	150.00	150.00	30.00	124	-	290
b	22.50	270.00	270.00	115.71	142	-	86
c	112.50	200.00	50.00	57.14	126	-	255
d	22.50	360.00	90.00	240.00	142	-	84

NEEDED TO DETECT:			R	h_N^2	h_B^2
H_R	D_R''	D_R'			
-	-	-	508	-	-
-	-	-	137	-	-
-	-	-	440	-	-
-	-	-	117	-	-
-	-	-	∞	0.53	-
-	-	-	9788	0.92	-
-	-	-	25339	0.57	-
-	-	-	1633	0.94	-
-	-	-	∞	0.53	-
-	-	-	9646	0.92	-
-	-	-	31059	0.55	-
-	-	-	1589	0.93	-
∞	-	-	∞	0.50	0.53
14413^x	-	-	14266	1.12	0.92
28605^x	-	-	44801	0.74	0.56
854^x	-	-	5148	1.61	0.94
-	∞	∞	∞	0.53	-
-	∞	19183	14960	0.92	-
-	∞	66275	32335	0.56	-
-	68734	2730	2264	0.93	-

TABLE 5B.2

MODEL	TRUE PARAMETER VALUES				SAMPLE SIZES		
	E_1	D_R	H_R	M	E_1	E_2	D_R
6a	112.50	150.00	150.00	30.00	126	-	935
b	22.50	270.00	270.00	115.71	142	-	174
c	112.50	200.00	50.00	57.14	128	-	1107
d	22.50	360.00	90.00	240.00	144	-	196
7a	112.50	150.00	150.00	30.00	194	∞	4670
b	22.50	270.00	270.00	115.71	279	∞	546
c	112.50	200.00	50.00	57.14	199	∞	2006
d	22.50	360.00	90.00	240.00	280	∞	211
8a	112.50	150.00	150.00	30.00	242	∞	255
b	22.50	270.00	270.00	115.71	285	11721	39
c	112.50	200.00	50.00	57.14	247	31126	208
d	22.50	360.00	90.00	240.00	286	2343	39
9a	112.50	150.00	150.00	30.00	199	∞	956
b	22.50	270.00	270.00	115.71	280	∞	174
c	112.50	200.00	50.00	57.14	202	∞	1128
d	22.50	360.00	90.00	240.00	281	∞	196
10a*	112.50	150.00	150.00	30.00	135	-	45299
b*	22.50	270.00	270.00	115.71	144	-	23333
c*	112.50	200.00	50.00	57.14	136	-	30348
d*	22.50	360.00	90.00	240.00	144	-	23054

NEEDED TO DETECT:			R	h_N^2	h_B^2
H_R	I	M			
-	-	∞	∞	0.53	-
-	-	6729	73152	0.92	-
-	-	14779	∞	0.56	-
-	-	747	∞	0.94	-
∞	-	-	∞	0.49	0.53
14849^x	-	-	13933	1.11	0.92
36882^x	-	-	53920	0.71	0.55
865^x	-	-	5006	1.61	0.94
-	∞^x	-	∞	0.53	0.51
-	12174^x	-	1976	0.92	0.83
-	53037^x	-	46728	0.57	0.50
-	2416^x	-	2552	0.94	0.74
-	-	∞	∞	0.53	-
-	-	6861	70488	0.92	-
-	-	18131	∞	0.56	-
-	-	754	∞	0.94	-
∞	-	91133	-	0.38	0.53
45341	-	10874	-	0.68	0.92
∞	-	30560	-	0.51	0.56
∞	-	4643	-	0.87	0.93

TABLE 6A.1

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	H_R	M
1	E_1	1	-	0.25	0.15	0.16
	E_2	-	1	0.25	0.10	0.38
2	E_1	1	0.46	-	0.01	-0.03
	D_R	-	1.04	1	0.49	1.14
3	E_1	1	-	-	0.01	-0.05
	E_2	-	1	-	-0.01	0.09
	D_R	-	-	1	0.51	0.95
4	E_1	1	0.50	-	-	0.00
	D_R	-	3	1	-	3.19
	H_R	-	-4	-	1	-4.17
5	E_1	1	-	-	-	-0.01
	E_2	-	1	-	-	0.05
	D_R	-	-	1	-	2.97
	H_R	-	-	-	1	-3.94
6	E_1	1	-	-	0.01	-0.01
	E_2	-	1	-	-0.04	0.27
	D_R	-	-	1	0.48	1.18
	I	-	-	-	0.09	-0.61
7	E_1	1	0.49	-	-	-
	D_R	-	-4.12	1	-	-
	H_R	-	5.33	-	1	-
	M	-	2.24	-	-	1
8	E_1	1	-	-	-	0.00
	E_2	-	1	-	-	0.14
	D_R	-	-	1	-	2.83
	H_R	-	-	-	1	-3.45
	I	-	-	-	-	-0.29

TABLE 6A.2

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	H_R	M
9	E_1	1	-	-	0.00	-
	E_2	-	1	-	0.02	-
	D_R	-	-	1	0.72	-
	I	-	-	-	-0.03	-
	M	-	-	-	-0.20	1
10	E_1	1	0.51	-	0.00	-
	D_R	-	-0.49	1	0.68	-
	M	-	1.33	-	-0.17	1
11	E_1	1	-	-	0.00	-
	E_2	-	1	-	0.01	-
	D_R	-	-	1	0.69	-
	M	-	-	-	-0.19	1
12	E_1	1	0.52	-	0.01	-0.03
	D_R	-	0.81	1	0.49	1.03
	D_R''	-	-0.47	-	0.02	0.07
	D_R'	-	0.85	-	-0.02	0.16
13	E_1	1	-	-	0.00	-0.02
	E_2	-	1	-	0.01	0.00
	D_R	-	-	1	0.48	1.03
	D_R''	-	-	-	0.03	-0.01
	D_R'	-	-	-	-0.03	0.17
14	E_1	1	0.53	-	0.00	-
	D_R	-	0.48	1	0.68	-
	M	-	0.33	-	-0.18	1
	D_R''	-	-0.52	-	0.03	-
	D_R'	-	0.80	-	0.01	-

TABLE 6B.1

MODEL	TRUE PARAMETER VALUES					SAM
	E_1	E_2	D_R	H_R	M	E_1
1a	56.25	56.25	150.00	150.00	30.00	91
b	56.25	56.25	200.00	50.00	57.14	91
c	11.25	11.25	270.00	270.00	115.71	91
d	11.25	11.25	360.00	90.00	240.00	92
2a	56.25	56.25	150.00	150.00	30.00	128
b	56.25	56.25	200.00	50.00	57.14	130
c	11.25	11.25	270.00	270.00	115.71	142
d	11.25	11.25	360.00	90.00	240.00	142
3a	56.25	56.25	150.00	150.00	30.00	227
b	56.25	56.25	200.00	50.00	57.14	232
c	11.25	11.25	270.00	270.00	115.71	284
d	11.25	11.25	360.00	90.00	240.00	285
4a	56.25	56.25	150.00	150.00	30.00	139
b	56.25	56.25	200.00	50.00	57.14	140
c	11.25	11.25	270.00	270.00	115.71	144
d	11.25	11.25	360.00	90.00	240.00	144
5a	56.25	56.25	150.00	150.00	30.00	253
b	56.25	56.25	200.00	50.00	57.14	255
c	11.25	11.25	270.00	270.00	115.71	285
d	11.15	11.25	360.00	90.00	240.00	285

PLE SIZES NEEDED TO DETECT:			R	h_N^2	h_B^2
E_2	D_R	H_R			
160	-	-	401	-	-
147	-	-	342	-	-
144	-	-	127	-	-
120	-	-	108	-	-
-	96	-	1095	0.66	-
-	87	-	977	0.68	-
-	37	-	1677	0.94	-
-	36	-	860	0.95	-
658	181	-	∞	0.53	-
616	163	-	26913	0.55	-
1291	39	-	9739	0.92	-
1234	37	-	1540	0.93	-
-	1154	19256^x	1095	0.86	0.65
-	667	4409^x	1056	1.08	0.67
-	468	9346^x	1789	1.18	0.94
-	190	737^x	1447	1.66	0.95
674	3475	∞	∞	0.50	0.53
669	1474	22190^x	71793	0.73	0.55
1314	522	13711^x	14774	1.12	0.92
1294	202	813	5240	1.61	0.94

TABLE 6B.2

MODEL	TRUE PARAMETER VALUES					SAMPLE SIZES NEEDED TO DETECT:						R	h_N^2	h_B^2
	E_1	E_2	D_R	H_R	M	E_1	E_2	D_R	H_R	I	M			
6a	56.25	56.25	150.00	150.00	30.00	270	1541	260	-	∞ x	-	∞	0.53	0.52
b	56.25	56.25	200.00	50.00	57.14	272	1190	213	-	46023 x	-	39218	0.57	0.51
c	11.25	11.25	270.00	270.00	115.71	286	5688	41	-	12216 x	-	19746	0.92	0.84
d	11.25	11.25	360.00	90.00	240.00	286	1648	39	-	2269 x	-	2454	0.94	0.74
7a	56.25	56.25	150.00	150.00	30.00	139	-	∞ x	13492	-	4874	1281	0.18	0.64
b	56.25	56.25	200.00	50.00	57.14	140	-	∞ x	24197	-	4157	1291	0.30	0.66
c	11.25	11.25	270.00	270.00	115.71	144	-	36775	30414	-	7740	2112	0.64	0.94
d	11.25	11.25	360.00	90.00	240.00	144	-	32614	∞	-	3960	2112	0.84	0.95
8a	56.25	56.25	150.00	150.00	30.00	278	1762	3801	∞	∞ x	-	∞	0.49	0.52
b	56.25	56.25	200.00	50.00	57.14	278	1732	1639	37755 x	∞ x	-	92737	0.71	0.53
c	11.25	11.25	270.00	270.00	115.71	286	9759	683	42360 x	25094 x	-	25051	1.05	0.86
d	11.25	11.25	360.00	90.00	240.00	286	6090	272	1503 x	11107 x	-	8585	1.50	0.85
9a*	56.25	56.25	150.00	150.00	30.00	255	817	62001	∞	-	∞	-	0.38	0.53
b*	56.25	56.25	200.00	50.00	57.14	256	831	40438	∞	-	48123	-	0.51	0.56
c*	11.25	11.25	270.00	270.00	115.71	285	1397	24480	47107	-	11619	-	0.68	0.92
d*	11.25	11.25	360.00	90.00	240.00	285	1397	23916	∞	-	4872	-	0.87	0.93
10a	56.26	56.25	150.00	150.00	30.00	274	2557	1295	-	∞ x	∞ x	∞	0.53	0.52
b	56.25	56.25	200.00	50.00	57.14	275	3041	1565	-	∞ x	23495	∞	0.56	0.55
c	11.25	11.25	270.00	270.00	115.71	286	42039	276	-	∞ x	15022	80955	0.92	0.90
d	11.25	11.25	360.00	90.00	240.00	286	98327	322	-	∞ x	1132	∞	0.94	0.93

TABLE 6B.3

MODEL	TRUE PARAMETER VALUES					SA
	E_1	E_2	D_R	H_R	M	E_1
11a	56.25	56.25	150.00	150.00	30.00	133
b	56.25	56.25	200.00	50.00	57.14	134
c	11.25	11.25	270.00	270.00	115.71	144
d	11.25	11.25	360.00	90.00	240.00	144
12a	56.25	56.25	150.00	150.00	30.00	254
b	56.25	56.25	200.00	50.00	57.14	255
c	11.25	11.25	270.00	270.00	115.71	285
d	11.25	11.25	360.00	90.00	240.00	285
13a	56.25	56.25	150.00	150.00	30.00	129
b	56.25	56.25	200.00	50.00	57.14	130
c	11.25	11.25	270.00	270.00	115.71	142
d	11.25	11.25	360.00	90.00	240.00	142
14a	56.25	56.25	150.00	150.00	30.00	276
b	56.25	56.25	200.00	50.00	57.14	276
c	11.25	11.25	270.00	270.00	115.71	286
d	11.25	11.25	360.00	90.00	240.00	286
15a	56.25	56.25	150.00	150.00	30.00	130
b	56.25	56.25	200.00	50.00	57.14	135
c	11.25	11.25	270.00	270.00	115.71	144
d	11.25	11.25	360.00	90.00	240.00	144

MPLE SIZES NEEDED TO DETECT: R h_N^2

E_2	D_R	M	D_R	D_R		
-	886	4580	-	-	1271	0.64
-	1016	182	-	-	1315	0.66
-	162	4594	-	-	2112	0.94
-	180	645	-	-	2190	0.95
752	652	∞	-	-	∞	0.53
785	756	15156	-	-	∞	0.56
1369	160	6838	-	-	72839	0.92
1389	178	720	-	-	∞	0.93
-	195	-	55151 x	1264	1833	0.64
-	178	-	62491 x	1070	1682	0.66
-	81	-	∞	6789	1848	0.94
-	80	-	∞	1796	1066	0.95
1103	345	-	∞	∞ x	∞	0.53
1128	300	-	∞	83606	30257	0.56
1406	87	-	∞	18709	14642	0.92
1437	84	-	81073	2586	2097	0.93
-	754	83246	41673 x	1987	1799	0.63
-	848	10200	29555 x	2191	1853	0.66
-	235	11182	∞ x	53305	1980	0.94
-	278	1140	∞ x	77210	2048	0.95

TABLE 7A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	I
1	E_1	1	-	0.25	0.25
	E_2	-	1	0.25	0.25
2	E_1	1	0.45	-	0.25
	D_R	-	1.07	1	0.25
3	E_1	1	-	-	0.07
	E_2	-	1	-	0.25
	D_R	-	-	1	0.43
4	E_1	1	0.50	-	0.25
	D_R	-	3	1	0.50
	H_R	-	-4	-	-
5	E_1	1	0.53	-	0.26
	D_R	-	0.77	1	0.40
	D_R''	-	-0.44	-	0.10
	D_R'	-	0.84	-	0.08
6	E_1	1	0.51	-	0.25
	D_R	-	-0.53	1	0.50
	M	-	1.36	-	-
7	E_1	1	-	-	0.05
	E_2	-	1	-	0.27
	D_R	-	-	1	-0.40
	H_R	-	-	-	1.62
8	E_1	1	-	-	0.04
	E_2	-	1	-	0.30
	D_R	-	-	1	0.91
	M	-	-	-	-0.50
9	E_1	1	0.48	-	0.25
	D_R	-	-4.07	1	0.50
	H_R	-	5.27	-	-
	M	-	2.23	-	-

TABLE 7B.1

MODEL	TRUE PARAMETER VALUES				SAMPLE	
	E_1	E_2	D_R	I	E_1	E_2
1a	56.25	56.25	180.00	45.00	91	160
b	11.25	11.25	360.00	45.00	91	160
c	56.25	56.25	135.00	90.00	91	160
d	11.25	11.25	315.00	90.00	91	160
2a	56.25	56.25	180.00	45.00	123	-
b	11.25	11.25	360.00	45.00	141	-
c	56.25	56.25	135.00	90.00	120	-
d	11.25	11.25	315.00	90.00	139	-
3a	56.25	56.25	180.00	45.00	215	477
b	11.25	11.25	360.00	45.00	281	489
c	56.25	56.25	135.00	90.00	213	395
d	11.25	11.25	315.00	90.00	282	354
4a	56.25	56.25	180.00	45.00	136	-
b	11.25	11.25	360.00	45.00	142	-
c	56.25	56.25	135.00	90.00	135	-
d	11.25	11.25	315.00	90.00	142	-
5a	56.25	56.25	180.00	45.00	124	-
b	11.25	11.25	360.00	45.00	141	-
c	56.25	56.25	135.00	90.00	121	-
d	11.25	11.25	315.00	90.00	140	-

SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
D_R	H_R	D_R''	D_R'			
-	-	-	-	215	-	-
-	-	-	-	152	-	-
-	-	-	-	665	-	-
-	-	-	-	171	-	-
120	-	-	-	802	0.59	-
44	-	-	-	689	0.88	-
159	-	-	-	636	0.54	-
51	-	-	-	479	0.83	-
269	-	-	-	31323	0.44	-
55	-	-	-	17844	0.81	-
352	-	-	-	8252	0.39	-
68	-	-	-	4805	0.74	-
1308	14413^{x}	-	-	807	0.83	0.58
868	∞^{x}	-	-	670	0.93	0.88
1545	15398^{x}	-	-	636	0.76	0.53
1011	∞^{x}	-	-	464	0.88	0.83
262	-	55006^{x}	1023	1252	0.55	-
96	-	∞^{x}	14785	665	0.87	-
336	-	∞^{x}	920	944	0.49	-
110	-	∞^{x}	9286	464	0.82	-

TABLE 7B.2

MODEL	TRUE PARAMETER VALUES				SAMPLE	
	E_1	E_2	D_R	I	E_1	E_2
6a	56.25	56.25	180.00	45.00	128	-
b	11.25	11.25	360.00	45.00	142	-
c	56.25	56.25	135.00	90.00	126	-
d	11.25	11.25	315.00	90.00	141	-
7a	56.25	56.25	180.00	45.00	247	487
b	11.25	11.25	360.00	45.00	284	488
c	56.25	56.25	135.00	90.00	248	397
d	11.25	11.25	315.00	90.00	285	354
8a*	56.25	56.25	180.00	45.00	269	1459
b*	11.25	11.25	360.00	45.00	286	37714
c*	56.25	56.25	135.00	90.00	269	1456
d*	11.25	11.25	315.00	90.00	286	37702
9a	56.25	56.25	180.00	45.00	248	524
b	11.25	11.25	360.00	45.00	284	493
c	56.25	56.25	135.00	90.00	249	418
d	11.25	11.25	315.00	90.00	285	357
10a	56.25	56.25	180.00	45.00	136	-
b	11.25	11.25	360.00	45.00	142	-
c	56.25	56.25	135.00	90.00	135	-
d	11.25	11.25	315.00	90.00	142	-

SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
D_R	H_R	I	M			
1472	-	-	4637	875	0.56	-
195	-	-	91307	674	0.87	-
2141	-	-	4726	680	0.51	-
243	-	-	92973	469	0.82	-
6704	∞	-	-	3510	0.36	0.44
1471	53345	-	-	21540	0.70	0.81
16347	33355	-	-	9069	0.23	0.39
2744	13703	-	-	5906	0.52	0.74
559	-	18666	-	-	0.40	0.50
66	-	10628	-	-	0.80	0.90
1109	-	5029	-	-	0.30	0.50
108	-	2958	-	-	0.70	0.90
808	-	-	61755^x	41878	0.44	-
174	-	-	22883^x	30609	0.81	-
865	-	-	17630^x	10501	0.38	-
185	-	-	6254^x	57113	0.73	-
∞^x	26501	-	6157	846	0.23	0.57
7949_x	∞	-	∞	645	0.80	0.87
∞^x	27475	-	6175	650	0.18	0.52
9078	∞	-	∞	450	0.75	0.82

TABLE 8A.1

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	I	M
1	E_1	1	-	0.25	0.25	0.16
	E_2	-	1	0.25	0.25	0.38
2	E_1	1	0.46	-	0.25	-0.04
	D_R	-	1.05	1	0.50	1.15
3	E_1	1	-	-	0.06	-0.05
	E_2	-	1	-	0.27	0.10
	D_R	-	-	1	0.43	0.93
4	E_1	1	0.50	-	0.25	0.01
	D_R	-	3	1	0.50	3.21
	H_R	-	-4	-	-	-4.22
5	E_1	1	-	-	0.40	-0.01
	E_2	-	1	-	0.29	0.53
	D_R	-	-	1	-0.44	2.97
	H_R	-	-	-	1.69	-3.95
6	E_1	1	-	-	-	-0.01
	E_2	-	1	-	-	0.26
	D_R	-	-	1	-	1.19
	I	-	-	-	1	-0.61
7	E_1	1	0.49	-	0.25	-
	D_R	-	-4.17	1	0.50	-
	H_R	-	5.43	-	-	-
	M	-	2.24	-	-	1
8	E_1	1	-	-	-	0.00
	E_2	-	1	-	-	0.14
	D_R	-	-	1	-	2.83
	H_R	-	-	-	-	-3.46
	I	-	-	-	1	-0.28

TABLE 8A.2

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	I	M
9	E_1	1	-	-	0.03	-
	E_2	-	1	-	0.33	-
	D_R	-	-	1	1.88	-
	H_R	-	-	-	-1.41	-
	M	-	-	-	-0.78	1
10	E_1	1	0.51	-	0.25	-
	D_R	-	-0.45	1	0.50	-
	M	-	1.28	-	-	1
11	E_1	1	-	-	0.03	-
	E_2	-	1	-	0.32	-
	D_R	-	-	1	0.90	-
	M	-	-	-	0.51	1
12	E_1	1	0.52	-	0.25	-0.04
	D_R	-	0.80	1	0.41	1.04
	D_R''	-	-0.45	-	0.08	0.07
	D_R'	-	0.85	-	0.08	0.16
13	E_1	1	-	-	0.00	-0.02
	E_2	-	1	-	0.45	0.00
	D_R	-	-	1	0.07	1.05
	D_R''	-	-	-	0.33	-0.03
	D_R'	-	-	-	-0.30	0.16
14	E_1	1	0.54	-	0.25	-
	D_R	-	0.50	1	0.52	-
	M	-	0.32	-	-0.09	1
	D_R''	-	-0.55	-	0.07	-
	D_R'	-	0.80	-	0.10	-

TABLE 8B.1.

MODEL	TRUE PARAMETER VALUES					SAM
	E_1	E_2	D_R	I	M	E_1
1a	56.25	56.25	180.00	45.00	45.00	91
b	56.25	56.25	135.00	90.00	23.82	91
c	11.25	11.25	360.00	45.00	240.00	92
d	11.25	11.25	315.00	90.00	169.62	92
2a	56.25	56.25	180.00	45.00	45.00	127
b	56.25	56.25	135.00	90.00	23.82	122
c	11.25	11.25	360.00	45.00	240.00	142
d	11.25	11.25	315.00	90.00	169.62	141
3a	56.25	56.25	180.00	45.00	45.00	229
b	56.25	56.25	135.00	90.00	23.82	220
c	11.25	11.25	360.00	45.00	240.00	285
d	11.25	11.25	315.00	90.00	169.62	285
4a	56.25	56.25	180.00	45.00	45.00	138
b	56.25	56.25	135.00	90.00	23.82	135
c	11.25	11.25	360.00	45.00	240.00	144
d	11.25	11.25	315.00	90.00	169.62	142
5a	56.25	56.25	180.00	45.00	45.00	253
b	56.25	56.25	135.00	90.00	23.82	251
c	11.25	11.25	360.00	45.00	240.00	285
d	11.25	11.25	315.00	90.00	169.62	285

IPLE SIZES NEEDED TO DETECT:			R	h_N^2	h_B^2
E_2	D_R	H_R			
147	-	-	418	-	-
153	-	-	564	-	-
117	-	-	113	-	-
124	-	-	131	-	-
-	107	-	749	0.63	-
-	140	-	623	0.57	-
-	38	-	442	0.92	-
-	44	-	381	0.88	-
464	213	-	36593	0.48	-
388	299	-	11227	0.41	-
482	45	-	1770	0.87	-
358	59	-	4038	0.79	-
-	761	4576^x	795	1.03	0.62
-	1111	7732^x	637	0.89	0.55
-	183	626^x	585	1.69	0.92
-	262	1188^x	439	1.49	0.88
497	2290	66744^x	50540	0.59	0.48
401	6436	∞	10969	0.36	0.41
486	221	839^x	8221	1.53	0.87
358	395	2608^x	12024	1.20	0.80

TABLE 8B.2

MODEL	TRUE PARAMETER VALUES					SAMPLE SIZES NEEDED TO DETECT:						R	h_N^2	h_B^2
	E_1	E_2	D_R	I	M	E_1	E_2	D_R	H_R	I	M			
6a	56.25	56.25	180.00	45.00	45.00	271	1174	370	-	∞	-	42973	0.47	0.50
b	56.25	56.25	135.00	90.00	23.82	270	1290	804	-	7623	-	∞	0.34	0.50
c	11.25	11.25	360.00	45.00	240.00	286	1464	49	-	4157 x	-	2163	0.88	0.73
d	11.25	11.25	315.00	90.00	169.62	286	2094	69	-	∞ x	-	3901	0.79	0.76
7a	56.25	56.25	180.00	45.00	45.00	138	-	∞ x	31736	-	4449	885	0.29	0.61
b	56.25	56.25	135.00	90.00	23.82	137	-	∞ x	30221	-	5103	665	0.22	0.54
c	11.25	11.25	360.00	45.00	240.00	142	-	28246	∞	-	3958	649	0.87	0.92
d	11.25	11.25	315.00	90.00	169.62	142	-	24001	∞	-	5350	452	0.82	0.87
8a	56.25	56.25	180.00	45.00	45.00	278	1748	2221	34925 x	49761	-	∞	0.62	0.52
b	56.25	56.25	135.00	90.00	23.82	278	1775	4956	∞ x	7676 x	-	∞	0.42	0.51
c	11.25	11.25	360.00	45.00	240.00	286	6048	278	1289 x	∞ x	-	8649	1.49	0.84
d	11.25	11.25	315.00	90.00	169.62	286	7523	432	2633 x	37488	-	14286	1.26	0.85
9a	56.25	56.25	180.00	45.00	45.00	279	584	21785	∞ x	-	∞	49159	0.55	0.48
b	56.25	56.25	135.00	90.00	23.82	254	457	16037	∞ x	-	83342 x	11372	0.52	0.41
c	11.25	11.25	360.00	45.00	240.00	284	510	13492	∞ x	-	7894	59452	0.96	0.87
d	11.25	11.25	315.00	90.00	169.62	284	377	7868	96952 x	-	41873	14060	0.97	0.80
10a	56.25	56.25	180.00	45.00	45.00	276	3040	2878	-	32054	25303	-	0.45	0.55
b	56.25	56.25	135.00	90.00	23.82	276	2813	5257	-	8036	88731	-	0.34	0.53
c	11.25	11.25	360.00	45.00	240.00	286	∞	517	-	35010	1003	-	0.87	0.93
d	11.25	11.25	315.00	90.00	169.62	286	∞	779	-	8762	2005	-	0.55	0.92

TABLE 8B.

MODEL	TRUE PARAMETER VALUES				
	E_1	E_2	D_R	I	M
11a	56.25	56.25	180.00	45.00	45.00
b	56.25	56.25	135.00	90.00	23.82
c	11.25	11.25	360.00	45.00	240.00
d	11.25	11.25	315.00	90.00	169.62
12a	56.25	56.25	180.00	45.00	45.00
b	56.25	56.25	135.00	90.00	23.82
c	11.25	11.25	360.00	45.00	240.00
d	11.25	11.25	315.00	90.00	169.62
13a	56.25	56.25	180.00	45.00	45.00
b	56.25	56.25	135.00	90.00	23.82
c	11.25	11.25	360.00	45.00	240.00
d	11.25	11.25	315.00	90.00	169.62
14a	56.25	56.25	180.00	45.00	45.00
b	56.25	56.25	135.00	90.00	23.82
c	11.25	11.25	360.00	45.00	240.00
d	11.25	11.25	315.00	90.00	169.62
15a	56.25	56.25	180.00	45.00	45.00
b	56.25	56.25	135.00	90.00	23.82
c	11.25	11.25	360.00	45.00	240.00
d	11.25	11.25	315.00	90.00	169.62

SAMPLES SIZE NEEDED TO DETECT:						R	h_N^2
E_1	E_2	D_R	M	D_R''	D_R'		
132	-	1527	2261	-	-	917	0.60
127	-	2186	3135	-	-	695	0.54
142	-	231	562	-	-	677	0.92
142	-	285	997	-	-	471	0.87
253	556	868	70188	-	-	49336	0.49
251	432	901	86566[x]	-	-	11509	0.41
285	506	211	779	-	-	57113	0.87
285	371	225	2559	-	-	13470	0.80
127	-	218	-	80412[x]	955	1164	0.60
122	-	293	-	∞ [x]	900	921	0.52
142	-	86	-	∞	1491	481	0.92
141	-	97	-	∞	1879	398	0.87
275	758	451	-	∞	∞ [x]	47033	0.47
275	565	803	-	28648	11383[x]	51661	0.36
286	536	101	-	67476	4672	1978	0.87
286	403	130	-	49634	∞	3886	0.79
133	-	1157	11882	35382[x]	1740	1224	0.60
129	-	1545	30768	70865[x]	1376	914	0.52
142	-	339	∞	28432	1038	633	0.92
142	-	379	∞	13750	2077	444	0.87

TABLE 9A.1

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	H_R	I
1	E_1	1	-	0.25	0.15	0.25
	E_2	-	1	0.25	0.10	0.25
2	E_1	1	0.45	-	0.01	0.25
	D_R	-	1.07	1	0.48	0.50
3	E_1	1	-	-	0.03	0.18
	E_2	-	1	-	-0.01	0.18
	D_R	-	-	1	0.46	0.29
4	E_1	1	0.50	-	-	0.25
	D_R	-	3	1	-	0.50
	H_R	-	-4	-	1	-
5	E_1	1	-	-	-	0.05
	E_2	-	1	-	-	0.27
	D_R	-	-	1	-	-0.40
	H_R	-	-	-	1	1.64
6	E_1	1	-	-	0.01	-
	E_2	-	1	-	-0.04	-
	D_R	-	-	1	0.47	-
	I	-	-	-	0.10	1
7	E_1	1	0.48	-	-	0.25
	D_R	-	-4.01	1	-	0.50
	H_R	-	5.17	-	1	-
	M	-	2.23	-	-	-
8	E_1	1	-	-	-	0.04
	E_2	-	1	-	-	0.31
	D_R	-	-	1	-	1.76
	H_R	-	-	-	1	-1.22
	M	-	-	-	-	-0.74

TABLE 9A.2

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	H_R	I
9	E_1	1	-	-	0.00	-
	E_2	-	1	-	0.02	-
	D_R	-	-	1	0.71	-
	I	-	-	-	-0.03	1
	M	-	-	-	-0.20	-
10	E_1	1	0.52	-	0.01	0.25
	D_R	-	-0.57	1	0.66	0.50
	M	-	1.40	-	-0.16	-
11	E_1	1	-	-	0.00	0.04
	E_2	-	1	-	0.01	0.30
	D_R	-	-	1	0.68	0.92
	M	-	-	-	-0.18	-0.51
12	E_1	1	0.53	-	0.01	0.26
	D_R	-	0.77	1	0.48	0.40
	D_R''	-	-0.44	-	0.02	0.10
	D_R'	-	0.84	-	-0.02	0.08
13	E_1	1	-	-	0.00	0.01
	E_2	-	1	-	0.01	0.44
	D_R	-	-	1	0.47	0.08
	D_R''	-	-	-	0.04	0.33
	D_R'	-	-	-	-0.04	-0.29
14	E_1	1	0.55	-	0.01	0.25
	D_R	-	0.34	1	0.66	0.51
	M	-	0.40	-	-0.17	-0.11
	D_R''	-	-0.46	-	0.03	0.11
	D_R'	-	0.78	-	0.01	0.09

TABLE 9

MODEL	TRUE PARAMETER VALUES				
	E_1	E_2	D_R	H_R	I
1a	56.25	56.25	120.00	120.00	45.00
b	56.25	56.25	90.00	90.00	90.00
c	56.25	56.25	160.00	40.00	45.00
d	56.25	56.25	120.00	30.00	90.00
e	11.25	11.25	240.00	240.00	45.00
f	11.25	11.25	210.00	210.00	90.00
g	11.25	11.25	320.00	80.00	45.00
h	11.25	11.25	280.00	70.00	90.00
2a	56.25	56.25	120.00	120.00	45.00
b	56.25	56.25	90.00	90.00	90.00
c	56.25	56.25	160.00	40.00	45.00
d	56.25	56.25	120.00	30.00	90.00
e	11.25	11.25	240.00	240.00	45.00
f	11.25	11.25	210.00	210.00	90.00
g	11.25	11.25	320.00	80.00	45.00
h	11.25	11.25	280.00	70.00	90.00
3a	56.25	56.25	120.00	120.00	45.00
b	56.25	56.25	90.00	90.00	90.00
c	56.25	56.25	160.00	40.00	45.00
d	56.25	56.25	120.00	30.00	90.00
e	11.25	11.25	240.00	240.00	45.00
f	11.25	11.25	210.00	210.00	90.00
g	11.25	11.25	320.00	80.00	45.00
h	11.25	11.25	280.00	70.00	90.00

B.1

SAMPLE SIZE NEEDED			R	h_N^2
E_1	E_2	D_R		
91	170	-	568	-
91	168	-	689	-
91	164	-	552	-
91	163	-	623	-
90	178	-	163	-
90	176	-	182	-
91	166	-	156	-
91	166	-	175	-
123	-	133	838	0.58
120	-	163	655	0.53
123	-	130	816	0.58
120	-	160	643	0.54
141	-	44	694	0.87
139	-	51	490	0.82
141	-	44	609	0.88
139	-	51	485	0.82
211	501	269	16628	0.44
212	410	357	6737	0.39
214	485	269	25972	0.44
212	400	353	7762	0.39
282	504	55	5781	0.82
282	361	68	2856	0.75
282	497	55	12406	0.81
282	357	68	4084	0.74

TABLE 9B.2

MODEL	TRUE PARAMETER VALUES					SAM
	E_1	E_2	D_R	H_R	I	E_1
4a	56.25	56.25	120.00	120.00	45.00	136
b	56.25	56.25	90.00	90.00	90.00	135
c	56.25	56.25	160.00	40.00	45.00	136
d	56.25	56.25	120.00	30.00	90.00	135
e	11.25	11.25	240.00	240.00	45.00	142
f	11.25	11.25	210.00	210.00	90.00	142
g	11.25	11.25	320.00	80.00	45.00	142
h	11.25	11.25	280.00	70.00	90.00	142
5a	56.25	56.25	120.00	120.00	45.00	248
b	56.25	56.25	90.00	90.00	90.00	249
c	56.25	56.25	160.00	40.00	45.00	247
d	56.25	56.25	120.00	30.00	90.00	248
e	11.25	11.25	240.00	240.00	45.00	284
f	11.25	11.25	210.00	210.00	90.00	285
g	11.25	11.25	320.00	80.00	45.00	284
h	11.25	11.25	280.00	70.00	90.00	285
6a	56.25	56.25	120.00	120.00	45.00	268
b	56.25	56.25	90.00	90.00	90.00	269
c	56.25	56.25	160.00	40.00	45.00	269
d	56.25	56.25	120.00	30.00	90.00	269
e	11.25	11.25	240.00	240.00	45.00	286
f	11.25	11.25	210.00	210.00	90.00	286
g	11.25	11.25	320.00	80.00	45.00	286
h	11.25	11.25	280.00	70.00	90.00	286

IPLE SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
E_2	D_R	H_R	I			
-	1942	69803[x]	-	819	0.69	0.56
-	2099	44365[x]	-	643	0.68	0.53
-	1485	21709[x]	-	811	0.78	0.58
-	1704	20754[x]	-	638	0.74	0.53
-	1898	16024	-	673	0.66	0.88
-	2051	23422	-	468	0.64	0.83
-	1100	∞	-	671	0.84	0.88
-	1255	∞	-	467	0.80	0.83
493	17845	18319	-	35104	0.23	0.44
401	53109	12906	-	8930	0.13	0.39
489	8883	51400	-	35104	0.31	0.44
399	22637	22947	-	9070	0.20	0.39
492	4203	5206	-	22018	0.44	0.81
357	9627	3679	-	5977	0.29	0.74
489	1994	17020	-	21718	0.61	0.81
355	3890	7829	-	5933	0.44	0.74
1746	592	-	11892	95600	0.39	0.52
1667	1181	-	4180	∞	0.29	0.51
1547	570	-	15875	∞	0.40	0.51
1523	1132	-	4719	∞	0.30	0.50
∞	67	-	5074	19418	0.80	0.94
∞	109	-	2014	26768	0.69	0.94
76563	66	-	7984	∞	0.80	0.91
69540	109	-	2579	∞	0.70	0.91

TABLE 9B.3

MODEL	TRUE PARAMETER VALUES					SAMPLE SIZES	
	E_1	E_2	D_R	H_R	I	E_1	E_2
7a	56.25	56.25	120.00	120.00	45.00	137	-
b	56.25	56.25	90.00	90.00	90.00	136	-
c	56.25	56.25	160.00	40.00	45.00	137	-
d	56.25	56.25	120.00	30.00	90.00	136	-
e	11.25	11.25	240.00	240.00	45.00	142	-
f	11.25	11.25	210.00	210.00	90.00	142	-
g	11.25	11.25	320.00	80.00	45.00	142	-
h	11.25	11.25	280.00	70.00	90.00	142	-
8a*	56.25	56.25	120.00	120.00	45.00	277	1795
b*	56.25	56.25	90.00	90.00	90.00	277	1802
c*	56.25	56.25	160.00	40.00	45.00	277	1787
d*	56.25	56.25	120.00	30.00	90.00	277	1797
e*	11.25	11.25	240.00	240.00	45.00	286	47911
f*	11.25	11.25	210.00	210.00	90.00	286	48366
g*	11.25	11.25	320.00	80.00	45.00	286	46995
h*	11.25	11.25	280.00	70.00	90.00	286	47664
9a	56.25	56.25	120.00	120.00	45.00	251	563
b	56.25	56.25	90.00	90.00	90.00	252	449
c	56.25	56.25	160.00	40.00	45.00	251	559
d	56.25	56.25	120.00	30.00	90.00	252	446
e	11.25	11.25	240.00	240.00	45.00	284	506
f	11.25	11.25	210.00	210.00	90.00	284	370
g	11.25	11.25	320.00	80.00	45.00	284	504
h	11.25	11.25	280.00	70.00	90.00	284	368

NEEDED TO DETECT:				R	h_N^2	h_B^2
D_R	H_R	I	M			
∞ x	13866	-	6164	896	0.10	0.57
∞ x	16526	-	6181	656	0.08	0.52
∞ x	20848	-	6160	851	0.18	0.57
∞ x	22846	-	6177	651	0.15	0.52
19087	20240	-	∞	647	0.54	0.87
20349	25326	-	∞	450	0.52	0.82
10222	89373	-	∞	646	0.72	0.87
11478	∞	-	∞	448	0.68	0.82
13487	56484	22439	-	-	0.27	0.50
23978	∞	6127	-	-	0.20	0.50
7363	∞	22439	-	-	0.36	0.50
13207	∞	6127	-	-	0.27	0.50
3322	12508	13538	-	-	0.53	0.90
4425	17077	3836	-	-	0.47	0.90
1761	∞	13346	-	-	0.71	0.90
2369	∞	3807	-	-	0.62	0.90
29656	∞	-	∞ x	41473	0.36	0.44
20652	∞ x	-	31199 x	10334	0.39	0.38
20476	∞ x	-	∞ x	41473	0.45	0.44
16203	∞ x	-	31173 x	10337	0.45	0.38
7227	∞	-	30196 x	34448	0.66	0.81
5016	∞	-	8452 x	8698	0.71	0.73
4791	∞ x	-	30286 x	33650	0.84	0.81
3687	∞ x	-	8484 x	8595	0.87	0.73

TABLE 9B.4

MODEL	TRUE PARAMETER VALUES					SAM
	E_1	E_2	D_R	H_R	I	E_1
10a	56.25	56.25	120.00	120.00	45.00	272
b	56.25	56.25	90.00	90.00	90.00	273
c	56.25	56.25	160.00	40.00	45.00	272
d	56.25	56.25	120.00	30.00	90.00	273
e	11.25	11.25	240.00	240.00	45.00	286
f	11.25	11.25	210.00	210.00	90.00	286
g	11.25	11.25	320.00	80.00	45.00	286
h	11.25	11.25	280.00	70.00	90.00	286
11a	56.25	56.25	120.00	120.00	45.00	128
b	56.25	56.25	90.00	90.00	90.00	124
c	56.25	56.25	160.00	40.00	45.00	128
d	56.25	56.25	120.00	30.00	90.00	126
e	11.25	11.25	240.00	240.00	45.00	142
f	11.25	11.25	210.00	210.00	90.00	141
g	11.25	11.25	320.00	80.00	45.00	142
h	11.25	11.25	280.00	70.00	90.00	141
12a	56.25	56.25	120.00	120.00	45.00	249
b	56.25	56.25	90.00	90.00	90.00	249
c	56.25	56.25	160.00	40.00	45.00	248
d	56.25	56.25	120.00	30.00	90.00	249
e	11.25	11.25	240.00	240.00	45.00	285
f	11.25	11.25	210.00	210.00	90.00	285
g	11.25	11.25	320.00	80.00	45.00	285
h	11.25	11.25	280.00	70.00	90.00	285

MPLE SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
E_2	D_R	I	M			
2302	2082	31684	83719[x]	∞	0.40	0.49
2377	3922	7671	∞ [x]	∞	0.30	0.49
2415	2407	28683	∞ [x]	∞	0.40	0.50
2462	4547	7449	∞ [x]	∞	0.30	0.50
28630	358	40202	18859[x]	55713	0.80	0.87
33438	536	7378	26490[x]	72603	0.70	0.88
50751	397	24318	∞ [x]	∞	0.80	0.90
55611	606	6169	∞ [x]	∞	0.70	0.89
-	1243	-	7769	880	0.56	-
-	1847	-	6806	682	0.51	-
-	1387	-	5457	879	0.56	-
-	2032	-	5307	679	0.51	-
-	175	-	47237[x]	669	0.88	-
-	217	-	86178[x]	470	0.83	-
-	188	-	∞	673	0.87	-
-	233	-	∞	469	0.82	-
519	694	-	16079[x]	43405	0.43	-
416	767	-	9257[x]	10817	0.38	-
523	766	-	34963[x]	43275	0.44	-
418	829	-	13821[x]	10656	0.38	-
491	157	-	4541[x]	29923	0.81	-
354	169	-	2574[x]	9111	0.73	-
493	168	-	11088[x]	34875	0.81	-
355	180	-	4368[x]	8571	0.73	-

TABLE 9B.5

MODEL	TRUE PARAMETER VALUES					SAMPLE
	E_1	E_2	D_R	H_R	I	E_1
13a	56.25	56.25	120.00	120.00	45.00	125
b	56.25	56.25	90.00	90.00	90.00	120
c	56.25	56.25	160.00	40.00	45.00	125
d	56.25	56.25	120.00	30.00	90.00	120
e	11.25	11.25	240.00	240.00	45.00	141
f	11.25	11.25	210.00	210.00	90.00	140
g	11.25	11.25	320.00	80.00	45.00	141
h	11.25	11.25	280.00	70.00	90.00	140
14a	56.25	56.25	120.00	120.00	45.00	275
b	56.25	56.25	90.00	90.00	90.00	275
c	56.25	56.25	160.00	40.00	45.00	275
d	56.25	56.25	120.00	30.00	90.00	275
e	11.25	11.25	240.00	240.00	45.00	286
f	11.25	11.25	210.00	210.00	90.00	286
g	11.25	11.25	320.00	80.00	45.00	286
h	11.25	11.25	280.00	70.00	90.00	286
15a	56.25	56.25	120.00	120.00	45.00	130
b	56.25	56.25	90.00	90.00	90.00	127
c	56.25	56.25	160.00	40.00	45.00	130
d	56.25	56.25	120.00	30.00	90.00	127
e	11.25	11.25	240.00	240.00	45.00	142
f	11.25	11.25	210.00	210.00	90.00	141
g	11.25	11.25	320.00	80.00	45.00	142
h	11.25	11.25	280.00	70.00	90.00	141

SIZES NEEDED TO DETECT:

E_2	D_R	M	D_R''	D_R'	R	h_N^2
-	269	-	67652^x	1148	1268	0.55
-	346	-	∞^x	991	960	0.48
-	265	-	58736^x	1062	1262	0.55
-	340	-	∞^x	943	951	0.49
-	96	-	∞^x	86473	648	0.87
-	110	-	∞^x	24557	484	0.82
-	96	-	∞^x	23011	663	0.87
-	110	-	∞^x	12217	467	0.82
731	622	-	59149	16812^x	47777	0.40
550	1043	-	18111	6494^x	25206	0.31
740	606	-	83375	23104^x	∞	0.41
554	1011	-	21056	7506^x	31117	0.32
508	125	-	65972	5631^x	16538	0.80
374	173	-	20180	2501^x	12729	0.70
512	124	-	∞	9693^x	62007	0.80
379	171	-	27348	3375^x	22595	0.71
-	976	∞^x	71673^x	1477	1210	0.55
-	1352	∞^x	∞^x	1262	916	0.48
-	1059	∞	50566^x	1476	1360	0.55
-	1456	∞	∞^x	1261	914	0.49
-	230	15889^x	∞	17914	628	0.87
-	272	15107^x	∞	10468	445	0.82
-	257	83792^x	∞^x	18436	633	0.87
-	292	61664^x	∞^x	10507	440	0.82

TABLE 10A

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS			
		E_1	E_2	D_R	C
1	E_1	1	0.25	-	-0.02
	E_2	-	0.25	1	0.44
2	E_1	1	-	0.47	-0.03
	D_R	-	1	1.00	0.90
3	E_1	1	-	-	-0.04
	E_2	-	-	1	0.20
	D_R	-	1	-	0.43
4	E_1	1	-	0.50	-0.00
	D_R	-	1	3	2.73
	H_R	-	-	-4	-3.68
5	E_1	1	-	0.51	-0.00
	D_R	-	1	0.85	0.47
	D_R''	-	-	-0.46	0.35
	D_R'	-	-	0.85	0.41
6	E_1	1	-	-	-0.02
	E_2	-	-	1	0.15
	D_R	-	1	-	1.79
	H_R	-	-	-	-2.48
7	E_1	1	-	0.51	-0.02
	D_R	-	1	-0.37	0.41
	M	-	-	1.22	0.46
8	E_1	1	-	-	-0.02
	E_2	-	-	1	0.37
	D_R	-	1	-	0.73
	I	-	-	-	-0.60
9	E_1	1	-	-	-0.03
	E_2	-	-	1	0.18
	D_R	-	1	-	0.24
	M	-	-	-	0.23

TABLE 10A contd.

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	C
10	E_1	1	-	0.49	-
	D_R	-	1	-4.37	6.00
	H_R	-	-	5.74	-8.00
	M	-	-	2.26	-1.00

TABLE 10B.1

MODEL	TRUE PARAMETER VALUES				SAMPLE	
	E_1	E_2	D_R	C	E_1	E_2
1a	56.25	61.61	225.00	150.00	93	95
b	11.25	98.04	405.00	270.00	95	66
c	56.25	6.76	225.00	50.00	90	209
d	11.25	10.33	405.00	90.00	92	114
2a	56.25	61.61	225.00	150.00	133	-
b	11.25	98.04	405.00	270.00	140	-
c	56.25	6.76	225.00	50.00	132	-
d	11.25	10.33	405.00	90.00	142	-
3a	56.25	61.61	225.00	150.00	245	354
b	11.25	98.04	405.00	270.00	285	215
c	56.25	6.76	225.00	50.00	217	6416
d	11.25	10.33	405.00	90.00	284	1359
4a	56.25	61.61	225.00	150.00	140	-
b	11.25	98.04	405.00	270.00	142	-
c	56.25	6.76	225.00	50.00	140	-
d	11.25	10.33	405.00	90.00	144	-
5a	56.25	61.61	225.00	150.00	130	-
b	11.25	98.04	405.00	270.00	139	-
c	56.25	6.76	225.00	50.00	133	-
d	11.25	10.33	405.00	90.00	142	-
6a	56.25	61.61	225.00	150.00	263	415
b	11.25	98.04	405.00	270.00	286	233
c	56.25	6.76	225.00	50.00	235	8632
d	11.25	10.33	405.00	90.00	284	1412

SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
D_R	H_R	D_R''	D_R'			
-	-	-	-	308	-	-
-	-	-	-	152	-	-
-	-	-	-	301	-	-
-	-	-	-	136	-	-
74	-	-	-	353	0.72	-
45	-	-	-	162	0.86	-
79	-	-	-	2408	0.71	-
37	-	-	-	930	0.94	-
180	-	-	-	765	0.50	-
85	-	-	-	456	0.65	-
110	-	-	-	2967	0.67	-
39	-	-	-	1608	0.91	-
349	1297^x	-	-	420	1.38	0.71
119	644^x	-	-	190	1.65	0.86
836	9952^x	-	-	2785	0.97	0.70
373	3955^x	-	-	1073	1.29	0.94
177	-	58581	346	744	0.66	-
105	-	∞	251	249	0.84	-
168	-	86949	3311	4381	0.68	-
83	-	89165	1905	1271	0.93	-
930	5268^x	-	-	829	0.85	0.53
388	1685^x	-	-	554	1.14	0.69
1124	17943^x	-	-	3235	0.87	0.67
416	5232^x	-	-	1981	1.22	0.92

TABLE 10B.2

MODEL	TRUE PARAMETER VALUES				SAMPLE	
	E_1	E_2	D_R	C	E_1	E_2
7a	56.25	61.61	225.00	150.00	136	-
b	11.25	98.04	405.00	270.00	142	-
c	56.25	6.76	225.00	50.00	136	-
d	11.25	10.33	405.00	90.00	143	-
8a	56.25	61.61	225.00	150.00	274	473
b	11.25	98.04	405.00	270.00	286	275
c	56.25	6.76	225.00	50.00	269	9767
d	11.25	10.33	405.00	90.00	286	4460
9a	56.25	61.61	225.00	150.00	265	421
b	11.25	98.04	405.00	270.00	286	240
c	56.25	6.76	225.00	50.00	244	7614
d	11.25	10.33	405.00	90.00	284	1409
10a	56.25	61.61	225.00	150.00	141	-
b	11.25	98.04	405.00	270.00	142	-
c	56.25	6.76	225.00	50.00	140	-
d	11.25	10.33	405.00	90.00	143	-
11a	56.25	61.61	225.00	150.00	233	634
b	11.25	98.04	405.00	270.00	260	284
c	56.25	6.76	255.00	50.00	212	25332
d	11.25	10.33	405.00	90.00	283	1521

SIZES NEEDED TO DETECT:

D_R	H_R	I	M	C	R	h_N^2	h_B^2
767	-	-	1814	-	379	0.71	-
331	-	-	831	-	175	0.86	-
136	-	-	36445	-	2424	0.70	-
159	-	-	9114	-	960	0.94	-
213	-	5582^x	-	-	831	0.57	0.42
97	-	1417^x	-	-	584	0.73	0.50
120	-	69316^x	-	-	2964	0.67	0.63
42	-	8987^x	-	-	1769	0.92	0.82
686	-	-	34013	-	752	0.51	-
263	-	-	3238	-	491	0.68	-
445	-	-	∞	-	2860	0.67	-
158	-	-	16976	-	1658	0.92	-
1944	4055^x	-	∞ ^{x}	-	404	1.42	0.71
1169	2272^x	-	∞ ^{x}	-	184	1.77	0.86
2382	9585^x	-	36120^x	-	2840	1.66	0.70
1265	4642^x	-	16865^x	-	1082	1.80	0.94
247	-	-	-	631	-	0.30	0.69
100	-	-	-	461	-	0.35	0.81
163	-	-	-	2000	-	0.50	0.72
60	-	-	-	1232	-	0.64	0.93

TABLE 11A.1

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS			
		E_1	D_R	D_R''	D_R'
1	E_1	1	0.25	0.23	-0.52
	E_2	-	0.25	-0.01	0.74
2	E_1	1	-	0.00	-0.06
	D_R	-	1	0.55	0.54
3	E_1	1	-	0.04	-0.17
	E_2	-	-	0.11	0.89
	D_R	-	1	0.65	-1.10
4	E_1	1	-	0.00	-0.02
	D_R	-	1	0.65	3.43
	H_R	-	-	-0.19	-5.81
5	E_1	1	-	0.01	-0.02
	D_R	-	1	0.43	-1.00
	M	-	-	0.09	1.37
6	E_1	1	-	-0.07	-0.73
	E_2	-	-	0.19	1.00
	D_R	-	1	-0.06	-
	H_R	-	-	0.75	-
7	E_1	1	-	-0.01	-0.05
	E_2	-	-	-0.24	1.22
	D_R	-	1	0.45	0.45
	I	-	-	0.54	-2.32
8	E_1	1	-	0.03	-0.14
	E_2	-	-	-0.09	0.65
	D_R	-	1	0.71	-1.53
	M	-	-	-0.08	0.80
9	E_1	1	-	-	-0.03
	D_R	-	1	-1.00	-2.17
	H_R	-	-	2.00	1.63
	M	-	-	0.50	1.71

TABLE 11A.2

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	D_R	D_R''	$-D_R'$
1	E_1	1	0.25	-0.21	-0.96
	E_2	-	0.25	0.45	1.20
2	E_1	1	-	0.01	-0.23
	D_R	-	1	0.52	0.99
3	E_1	1	-	-0.22	-0.84
	E_2	-	-	0.24	0.82
	D_R	-	1	0.51	0.63
4	E_1	1	-	0.01	-0.06
	D_R	-	1	0.42	3.34
	H_R	-	-	0.24	-5.38
5	E_1	1	-	0.03	-0.08
	D_R	-	1	0.02	-1.10
	M	-	-	0.43	1.62
6	E_1	1	-	-0.22	-0.78
	E_2	-	-	0.24	0.81
	D_R	-	1	0.64	1.43
	H_R	-	-	-0.32	-1.80
7	E_1	1	-	-0.37	-1.17
	E_2	-	-	0.21	0.77
	D_R	-	1	0.43	0.45
	I	-	-	0.44	0.89
8	E_1	1	-	-0.30	-1.10
	E_2	-	-	0.29	1.06
	D_R	-	1	0.64	1.27
	M	-	-	-0.07	-0.58
9	E_1	1	-	-	-0.08
	D_R	-	1	-1.00	-1.29
	H_R	-	-	2.00	0.31
	M	-	-	0.50	1.65

TABLE 11B.1

POPULATION		TRUE PARAMETER VALUES			
h_N^2	ρ	E_1	D_R	D_R''	D_R'
0.5	1.0	112.50	225.00	225.00	225.00
0.5	1.0	112.50	225.00	112.50	159.10
0.5	0.5	112.50	225.00	225.00	112.50
0.5	0.5	112.50	225.00	112.50	79.55
0.5	-1.0	112.50	225.00	225.00	-225.00
0.5	-1.0	112.50	225.00	112.50	-159.10
0.5	-0.5	112.50	225.00	225.00	-112.50
0.5	-0.5	112.50	225.00	112.50	- 79.55
0.9	1.0	22.50	405.00	405.00	405.00
0.9	1.0	22.50	405.00	202.50	286.38
0.9	0.5	22.50	405.00	405.00	202.50
0.9	0.5	22.50	405.00	202.50	143.19
0.9	-1.0	22.50	405.00	405.00	-405.00
0.9	-1.0	22.50	405.00	202.50	-286.38
0.9	-0.5	22.50	405.00	405.00	-202.50
0.9	-0.5	22.50	405.00	202.50	-143.19

SAMPLE SIZE: R

E_1	E_2	
85	72	305
85	90	452
83	143	322
82	173	466
47	147^x	343
51	42^x	386
67	∞	328
68	∞	416
88	36	149
88	39	188
85	71	116
85	72	143
35	39^x	136
38	56^x	135
68	∞^x	121
69	41603	135

TABLE 11B.2

POPULATION		TRUE PARAMETER VALUES			
h_N^2	ρ	E_1	D_R	D_R''	D_R'
0.5	1.0	112.50	225.00	225.00	225.00
0.5	1.0	112.50	225.00	112.50	159.10
0.5	0.5	112.50	225.00	225.00	112.50
0.5	0.5	112.50	225.00	112.50	79.55
0.5	-1.0	112.50	225.00	225.00	-225.00
0.5	-1.0	112.50	225.00	112.50	-159.10
0.5	-0.5	112.50	225.00	225.00	-112.50
0.5	-0.5	112.50	225.00	112.50	- 79.55
0.9	1.0	22.50	405.00	405.00	405.00
0.9	1.0	22.50	405.00	202.50	286.38
0.9	0.5	22.50	405.00	405.00	202.50
0.9	0.5	22.50	405.00	202.50	143.19
0.9	-1.0	22.50	405.00	405.00	-405.00
0.9	-1.0	22.50	405.00	202.50	-286.38
0.9	-0.5	22.50	405.00	405.00	-202.50
0.9	-0.5	22.50	405.00	202.50	-143.19

SAMPLE SIZE:		R	h_N^2
E_1	D_R		
132	80	185	0.70
128	98	306	0.65
129	93	443	0.66
126	114	835	0.61
101	769	132	0.27
104	542	227	0.31
117	185	578	0.50
115	218	1091	0.47
142	36	82	0.94
142	37	119	0.93
142	37	207	0.94
142	38	342	0.93
142	39	51	0.91
141	42	77	0.89
142	38	205	0.92
142	39	338	0.91

TABLE 11B.3

POPULATION		TRUE PARAMETER VALUES			
h_N^2	ρ	E_1	D_R	D_R''	D_R'
0.5	1.0	112.50	225.00	225.00	225.00
0.5	1.0	112.50	225.00	112.50	159.10
0.5	0.5	112.50	225.00	225.00	112.50
0.5	0.5	112.50	225.00	112.50	79.55
0.5	-1.0	112.50	225.00	225.00	-225.00
0.5	-1.0	112.50	225.00	112.50	-159.10
0.5	-0.5	112.50	225.00	225.00	-112.50
0.5	-0.5	112.50	225.00	112.50	- 79.55
0.9	1.0	22.50	405.00	405.00	405.00
0.9	1.0	22.50	405.00	202.50	286.38
0.9	0.5	22.50	405.00	405.00	202.50
0.9	0.5	22.50	405.00	202.50	143.19
0.9	-1.0	22.50	405.00	405.00	-405.00
0.9	-1.0	22.50	405.00	202.50	-286.38
0.9	0.5	22.50	405.00	405.00	-202.50
0.9	0.5	22.50	405.00	202.50	-143.19

SAMPLE SIZE NEEDED: R h_N^2

E_1	E_2	D_R		
214	157	1136	286	0.19
208	275	621	476	0.29
208	1017	217	568	0.50
196	1416	248	1185	0.48
68	150[x]	298	513	0.45
78	220[x]	285	832	0.47
129	1146[x]	165	1193	0.60
128	1488[x]	200	3912	0.55
245	41	17341	143	0.01
267	51	734	188	0.09
281	8675	38	202	0.93
279	7816	41	335	0.91
37	42[x]	47	119	0.87
47	57[x]	47	195	0.87
266	1450[x]	38	215	0.93
261	2393[x]	39	365	0.92

TABLE

POPULATION		TRUE PARAMETER VALUES		
h_N^2	ρ	E_1	D_R	D_R''
0.5	1.0	112.50	225.00	225.00
0.5	1.0	112.50	225.00	112.50
0.5	0.5	112.50	225.00	225.00
0.5	0.5	112.50	225.00	112.50
0.5	-1.0	112.50	225.00	225.00
0.5	-1.0	112.50	225.00	112.50
0.5	-0.5	112.50	225.00	225.00
0.5	-0.5	112.50	225.00	112.50
0.9	1.0	22.50	405.00	405.00
0.9	1.0	22.50	405.00	202.50
0.9	-0.5	22.50	405.00	405.00
0.9	-0.5	22.50	405.00	202.50
0.9	1.0	22.50	405.00	405.00
0.9	1.0	22.50	405.00	202.50
0.9	-0.5	22.50	405.00	405.00
0.9	-0.5	22.50	405.00	202.50

11B.4

D_R'	E_1	D_R	H_R	SAMPLE SIZES NEEDED: R	h_N^2	h_B^2
225.00	140	214	524[x]	223	1.67	0.68
159.10	139	337	1002[x]	354	1.43	0.63
112.50	139	527	2392[x]	497	1.19	0.65
79.55	138	779	4431[x]	922	1.03	0.60
-225.00	130	1358[x]	741	146	-0.95	0.44
-159.10	129	3420[x]	1347	245	-0.60	0.42
-112.50	135	∞[x]	3060	661	-0.09	0.54
- 79.55	133	∞	1088	1231	0.03	0.50
405.00	144	71	104[x]	104	2.37	0.95
286.38	144	90	163[x]	146	2.17	0.94
202.50	144	160	504[x]	276	1.76	0.94
143.19	144	202	792[x]	446	1.63	0.93
-405.00	144	342[x]	184	54	-1.88	0.91
-286.38	144	790[x]	297	81	-1.24	0.89
-202.50	144	48622[x]	907	257	-0.15	0.92
-143.19	144	∞	1288	422	0.04	0.91

TABLE 11B.5

POPULATION		TRUE PARAMETER VALUES				SAMP
h_N^2	ρ	E_1	D_R	D_R''	D_R'	E_1
0.5	1.0	112.50	225.00	225.00	225.00	135
0.5	1.0	112.50	225.00	112.50	159.10	133
0.5	0.5	112.50	225.00	225.00	112.50	133
0.5	0.5	112.50	225.00	112.50	79.55	130
0.5	-1.0	112.50	225.00	225.00	-225.00	118
0.5	-1.0	112.50	225.00	112.50	-159.10	117
0.5	0.5	112.50	225.00	225.00	-112.50	126
0.5	0.5	112.50	225.00	112.50	- 79.55	123
0.9	1.0	22.50	405.00	405.00	405.00	144
0.9	1.0	22.50	405.00	202.50	286.38	144
0.9	0.5	22.50	405.00	405.00	202.50	144
0.9	0.5	22.50	405.00	202.50	143.19	142
0.9	1.0	22.50	405.00	405.00	-405.00	142
0.9	1.0	22.50	405.00	202.50	-286.38	142
0.9	-0.5	22.50	405.00	405.00	-202.50	142
0.9	-0.5	22.50	405.00	202.50	-143.19	142

LE SIZE NEEDED		R	h_N^2
D_R	M		
5969	395	243	0.66
4391	689	392	0.62
1738	1258	566	0.63
1776	2213	1064	0.59
189	251[x]	176	0.43
254	421[x]	230	0.43
415	2836[x]	680	0.54
499	3565[x]	1335	0.51
333	103	11	0.94
308	154	154	0.93
212	394	319	0.94
217	595	527	0.93
42	43[x]	74	0.91
49	60[x]	115	0.90
103	765[x]	273	0.93
108	825[x]	478	0.91

TABLE 11B.6

POPULATION		TRUE PARAMETER VALUES				SAM
h_N^2	ρ	E_1	D_R	D_R''	D_R'	E_1
0.5	1.0	112.50	225.00	225.00	225.00	118
0.5	1.0	112.50	225.00	112.50	159.10	120
0.5	0.5	112.50	225.00	225.00	112.50	124
0.5	0.5	112.50	225.00	112.50	79.55	123
0.5	-1.0	112.50	225.00	225.00	-225.00	117
0.5	-1.0	112.50	225.00	112.50	-159.10	118
0.5	-0.5	112.50	225.00	225.50	-112.50	126
0.5	-0.5	112.50	225.00	112.50	- 79.55	124
0.9	1.0	22.50	405.00	405.00	405.00	129
0.9	1.0	22.50	405.00	202.50	286.38	136
0.9	0.5	22.50	405.00	405.00	202.50	142
0.9	0.5	22.50	405.00	202.50	143.19	142
0.9	-1.0	22.50	405.00	405.00	-405.00	128
0.9	-1.0	22.50	405.00	202.50	-286.38	139
0.9	-0.5	22.50	405.00	405.00	-202.50	142
0.9	-0.5	22.50	405.00	202.50	-143.19	142

MPLE SIZES NEEDED TO DETECT: R h_N^2

D_R	D_R''	D_R'		
322	835	148	-	0.50
323	2513	194	-	0.50
326	778	539	-	0.50
326	2463	753	-	0.50
322	1319	272[x]	-	0.50
322	3743	351[x]	-	0.50
327	1020	783[x]	-	0.50
326	3080	1066[x]	-	0.50
90	487	99	-	0.90
90	1218	108	-	0.90
90	364	275	-	0.90
90	984	316	-	0.90
90	711	156[x]	-	0.90
90	1715	170[x]	-	0.90
90	475	388[x]	-	0.90
90	1233	440[x]	-	0.90

TABLE 11B.7

POPULATION		TRUE PARAMETER VALUES				SAM
h_N^2	ρ	E_1	D_R	D_R''	D_R'	E_1
0.5	1.0	112.50	225.00	225.00	225.00	251
0.5	1.0	112.50	225.00	112.50	159.10	112
0.5	0.5	112.50	225.00	225.00	112.50	232
0.5	0.5	112.50	225.00	112.50	79.55	221
0.5	-1.0	112.50	225.00	225.00	-225.00	81
0.5	-1.0	112.50	225.00	112.50	-159.10	93
0.5	-0.5	112.50	225.00	225.00	-112.50	159
0.5	-0.5	112.50	225.00	112.50	- 79.55	157
0.9	1.0	22.50	405.00	405.00	405.00	285
0.9	1.0	22.50	405.00	202.50	286.38	280
0.9	0.5	22.50	405.00	405.00	202.50	281
0.9	0.5	22.50	405.00	202.50	143.19	280
0.9	-1.0	22.50	405.00	405.00	-405.00	39
0.9	-1.0	22.50	405.00	202.50	-286.38	50
0.9	-0.5	22.50	405.00	405.00	-202.50	279
0.9	-0.5	22.50	405.00	202.50	-143.19	275

MPLE SIZE NEEDED TO DETECT:			R	h_N^2	h_B^2
E_2	D_R	H_R			
234	1430	3484^x	276	0.64	0.27
378	1313	3777^x	489	0.75	0.33
1334	1242	8696^x	578	0.82	0.52
1759_x	1714	14354^x	1215	0.73	0.49
164^x	79594	7558	505	0.11	0.49
243^x	27826	13822	811	0.20	0.50
1454^x	11400	16014	1158	0.33	0.60
1811^x	10144	26836	2622	0.35	0.56
42	38977	∞^x	137	0.03	0.01
266	99	126^x	150	1.98	0.72
16211	166	527^x	265	1.74	0.93
11072	211_x	832^x	435	1.60	0.92
45^x	237^x	91	138	-0.78	0.89
61^x	7673^x	454	198	-0.26	0.88
5653^x	∞^x	1018	256	-0.08	0.93
5323^x	83458	1488	430	0.11	0.92

TABLE 11B.8

POPULATION		TRUE PARAMETER VALUES				SAM
h_N^2	ρ	E_1	D_R	D_R''	D_R'	E_1
0.5	1.0	112.50	225.00	225.00	225.00	238
0.5	1.0	112.50	225.00	112.50	159.10	241
0.5	0.5	112.50	225.00	225.00	112.50	257
0.5	0.5	112.50	225.00	112.50	79.55	250
0.5	-1.0	112.50	225.00	225.00	-225.00	91
0.5	-1.0	112.50	225.00	112.50	-159.10	138
0.5	-0.5	112.50	225.00	225.00	-112.50	242
0.5	-0.5	112.50	225.00	112.50	- 79.55	233
0.9	1.0	22.50	405.00	405.00	405.00	241
0.9	1.0	22.50	405.00	202.50	286.38	263
0.9	0.5	22.50	405.00	405.00	202.50	285
0.9	0.5	22.50	405.00	202.50	143.19	284
0.9	-1.0	22.50	405.00	405.00	-405.00	284
0.9	-1.0	22.50	405.00	202.50	-286.38	284
0.9	-0.5	22.50	405.00	405.00	-202.50	285
0.9	-0.5	22.50	405.00	202.50	-143.19	285

IPLE SIZE NEEDED TO DETECT:			R	h_N^2	h_B^2
E_2	D_R	I			
122	152	321^x	386	0.64	0.04
187	193	546^x	714	0.59	0.08
848	168	2433^x	656	0.61	0.35
1136	213	3640^x	1433	0.57	0.34
197^x	317	3986^x	551	0.49	0.26
292^x	346	∞	787	0.48	0.49
1031^x	236	2866	467	0.55	0.82
1394^x	280	4784	3613	0.51	0.74
45	45	48^x	209	0.94	0.00
54	47	59^x	357	0.93	0.05
349	41	367^x	338	0.94	0.50
371	43	397^x	675	0.93	0.50
50^x	51	60^x	268	0.90	0.01
59^x	48	73	226	0.89	1.71
404^x	39	431	547	0.92	1.37
452^x	42	489	1190	0.91	1.33

TABLE 11B.9

POPULATION		TRUE PARAMETER VALUES			
h_N^2	ρ	E_1	D_R	D_R''	D_R'
0.5	1.0	112.50	225.00	225.00	225.00
0.5	1.0	112.50	225.00	112.50	159.10
0.5	0.5	112.50	225.00	225.00	112.50
0.5	0.5	112.50	225.00	112.50	79.55
0.5	-1.0	112.50	225.00	225.00	-225.00
0.5	-1.0	112.50	225.00	112.50	-159.10
0.5	-0.5	112.50	225.00	225.00	-112.50
0.5	-0.5	112.50	225.00	112.50	- 79.55
0.9	1.0	22.50	405.00	405.00	405.00
0.9	1.0	22.50	405.00	202.50	286.38
0.9	0.5	22.50	405.00	405.00	202.50
0.9	0.5	22.50	405.00	202.50	143.19
0.9	-1.0	22.50	405.00	405.00	-405.00
0.9	-1.0	22.50	405.00	202.50	-286.38
0.9	-0.5	22.50	405.00	405.00	-202.50
0.9	-0.5	22.50	405.00	405.00	-143.19

SAMPLE SIZES NEEDED TO DETECT: R h_N^2

E_1	E_2	D_R	M		
242	230	36918	1818	271	0.32
232	516	10469	2604	486	0.36
229	2387	1758	3123	606	0.54
48	63	260	224	1262	0.50
73	185^x	3604	1649	688	0.46
91	290^x	1778	5987	924	0.47
166	1485^x	719	∞	1155	0.59
165	2140^x	778	∞^x	2537	0.56
284	127	∞	225	95	0.40
278	369	1265	112	150	0.12
280	34712	213	407	307	0.93
280	7417	219	621	510	0.92
45	65^x	5741	117	202	0.69
77	96^x	250	554^x	221	0.85
281	8755^x	105	907^x	268	0.93
280	9432^x	111	993^x	475	0.92

TABLE 11B.10

POPULATION		TRUE PARAMETER VALUES			
h_N^2	ρ	E_1	D_R	D_R''	D_R'
0.5	1.0	112.50	225.00	225.00	225.00
0.5	1.0	112.50	225.00	112.50	159.10
0.5	0.5	112.50	225.00	225.00	112.50
0.5	0.5	112.50	225.00	112.50	79.55
0.5	-1.0	112.50	225.00	225.00	-225.00
0.5	-1.0	112.50	225.00	112.50	-159.10
0.5	-0.5	112.50	225.00	225.00	-112.50
0.5	-0.5	112.50	225.00	112.50	79.55
0.9	1.0	22.50	405.00	405.00	405.00
0.9	1.0	22.50	405.00	202.50	286.38
0.9	0.5	22.50	405.00	405.00	202.50
0.9	0.5	22.50	405.00	202.50	143.19
0.9	-1.0	22.50	405.00	405.00	-405.00
0.9	-1.0	22.50	405.00	202.50	-286.38
0.9	-0.5	22.50	405.00	405.00	-202.50
0.9	-0.5	22.50	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT:				R	h_N^2	h_B^2
E_1	D_R	H_R	M			
140	16205[x]	10910	1532	225	0.01	0.66
139	53021[x]	22159	2204	369	0.20	0.62
139	50996[x]	14062	2667	531	0.13	0.64
138	∞[x]	33277	4406	1012	0.13	0.59
133	4762	9589	327[x]	171	0.07	0.46
132	3622	74013	555[x]	287	0.29	0.44
135	29139	16872	12039[x]	670	0.20	0.55
134	13090	75773	9474[x]	1294	0.34	0.59
144	6477[x]	4695	947	98	-0.25	0.94
144	14017[x]	7448	1188	143	0.08	0.93
144	17855[x]	5925	1501	290	0.06	0.94
144	∞[x]	11042	2115	490	0.34	0.92
144	1734	882	45[x]	71	-0.19	0.92
144	856	4174	62[x]	108	0.37	0.90
144	7464	5815	3159[x]	268	0.38	0.92
144	3236	23064	2195[x]	462	0.64	0.91

TABLE 12 A1.1

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	D_R''	D_R'
1	E_1	1	-	0.25	0.23	-0.52
	E_2	-	1	0.25	0.00	0.75
2	E_1	1	0.48	-	0.00	-0.05
	D_R	-	0.99	1	0.55	0.49
3	E_1	1	-	-	0.01	-0.10
	E_2	-	1	-	-0.06	0.61
	D_R	-	-	1	0.63	-0.85
4	E_1	1	0.50	-	0.00	-0.01
	D_R	-	3	1	0.64	3.42
	H_R	-	-4	-	-0.17	-5.84
5	E_1	1	0.51	-	0.00	-0.01
	D_R	-	-0.25	1	0.43	-1.05
	M	-	1.10	-	0.10	1.39
6	E_1	1	0.52	-	-	-
	D_R	-	0.82	1	-	-
	D_R''	-	-0.44	-	1	-
	D_R'	-	0.86	-	-	1
7	E_1	1	-	-	0.01	-0.07
	E_2	-	1	-	-0.05	0.51
	D_R	-	-	1	0.18	1.69
	H_R	-	-	-	0.87	-4.64
8	E_1	1	-	-	-0.00	-0.02
	E_2	-	1	-	-0.20	1.20
	D_R	-	-	1	0.45	0.40
	I	-	-	-	0.46	-2.32
9	E_1	1	-	-	0.01	-0.07
	E_2	-	1	-	-0.06	0.44
	D_R	-	-	1	0.71	-1.70
	M	-	-	-	-0.08	1.15

TABLE 12A 1.2

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	D_R''	D_R'
10	E_1	1	0.49	-	-	-0.02
	D_R	-	-4.38	1	-1	-2.18
	H_R	-	5.75	-	2	1.58
	M	-	2.26	-	0.50	1.72
11	E_1	1	0.52	-	-	-
	D_R	-	0.74	1	-	-
	M	-	0.08	-	-	-
	D_R''	-	-0.45	-	1	-
	D_R'	-	0.85	-	-	1
12	E_1	1	-	-	-0.04	-0.60
	E_2	-	1	-	-0.20	1.17
	D_R	-	-	1	0.47	0.69
	H_R	-	-	-	-0.04	-0.60
	I	-	-	-	0.46	-2.26
13	E_1	1	-	-	0.01	-0.07
	E_2	-	1	-	-0.08	0.44
	D_R	-	-	1	-1.84	-0.91
	H_R	-	-	-	3.52	-1.10
	M	-	-	-	0.68	0.92
14	E_1	1	-	-	0.00	-0.02
	E_2	-	1	-	-0.26	1.21
	D_R	-	-	1	0.21	0.44
	I	-	-	-	0.57	-2.34
	M	-	-	-	0.21	-0.03

TABLE 12A 2.1

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	D_R''	D_R'
1	E_1	1	-	0.25	-0.10	-0.85
	E_2	-	1	0.25	0.36	1.11
2	E_1	1	0.44	-	0.00	-0.12
	D_R	-	1.10	1	0.54	0.73
3	E_1	1	-	-	-0.12	-0.66
	E_2	-	1	-	0.15	0.76
	D_R	-	-	1	0.50	0.40
4	E_1	1	0.50	-	0.00	-0.03
	D_R	-	3	1	0.46	3.45
	H_R	-	-4	-	0.18	-5.78
5	E_1	1	0.52	-	0.01	-0.04
	D_R	-	-1.09	1	0.15	-1.22
	M	-	1.88	-	0.35	1.62
6	E_1	1	0.56	-	-	-
	D_R	-	0.71	1	-	-
	D_R''	-	-0.58	-	1	-
	D_R'	-	0.88	-	-	1
7	E_1	1	-	-	-0.11	-0.58
	E_2	-	1	-	0.15	0.72
	D_R	-	-	1	0.52	1.43
	H_R	-	-	-	-0.07	-2.20
8	E_1	1	-	-	-0.01	-0.09
	E_2	-	1	-	0.17	0.91
	D_R	-	-	1	0.44	0.39
	I	-	-	-	-0.22	-1.57
9	E_1	1	-	-	-0.10	-0.68
	E_2	-	1	-	0.12	0.76
	D_R	-	-	1	0.31	0.36
	M	-	-	-	0.20	0.07

TABLE 12A 2.2

MODEL	ESTIMATED PARAMETERS	TRUE PARAMETERS E_1	E_2	D_R	D_R''	D_R'
10	E_1	1	0.49	-	-	-0.04
	D_R	-	-3.47	1	-1	-1.42
	H_R	-	4.10	-	2	0.29
	M	-	2.22	-	0.50	1.67
11	E_1	1	0.58	-	-	-
	D_R	-	0.04	1	-	-
	M	-	0.72	-	-	-
	D_R''	-	-0.62	-	1	-
	D_R'	-	0.74	-	-	1
12	E_1	1	-	-	-0.02	-0.06
	E_2	-	1	-	0.19	0.89
	D_R	-	-	1	0.33	1.89
	H_R	-	-	-	0.18	-3.16
	I	-	-	-	-0.24	-1.48
13	E_1	1	-	-	-0.12	-0.68
	E_2	-	1	-	0.14	0.79
	D_R	-	-	1	0.21	1.85
	H_R	-	-	-	0.22	-2.52
	M	-	-	-	0.17	-0.20
14	E_1	1	-	-	-0.01	-0.09
	E_2	-	1	-	-0.04	0.93
	D_R	-	-	1	-0.18	0.49
	I	-	-	-	0.27	-1.61
	M	-	-	-	0.50	-0.07

TABLE 12B.1

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE		R
E_1	E_2	
95	51	315
93	87	278
95	56	468
93	90	394
62	728[x]	272
81	1373	260
69	5084[x]	301
84	753	318
96	37	150
95	72	108
96	39	174
95	71	131
39	49[x]	132
75	57	118
44	73[x]	131
77	8598	130

TABLE 12B.2

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE		R	h_N^2
E_1	D_R		
135	62	124	0.77
134	68	249	0.75
133	71	172	0.74
132	78	341	0.71
120	109	201	0.53
128	98	797	0.65
120	160	404	0.54
126	110	1512	0.62
144	35	73	0.96
142	36	179	0.95
142	36	105	0.95
142	36	271	0.95
142	37	53	0.93
142	36	203	0.94
142	38	81	0.92
142	37	332	0.93

TABLE 12B.3

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R	D_R
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	22.500	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE			R	h_N^2
E_1	E_2	D_R		
236	63	4349	299	0.06
250	268	206	566	0.46
238	84	1418	474	0.13
244	276	238	1151	0.43
101	439[x]	196	419	0.56
41	51	124	744	0.65
126	1284[x]	196	598	0.56
196	7329	146	1453	0.61
243	38	66275	144	0.00
285	926	38	197	0.93
275	44	509	191	0.07
285	893	39	325	0.91
42	49[x]	47	126	0.88
284	3752	37	197	0.94
55	68[x]	45	197	0.88
282	3518	38	326	0.93

TABLE 12B.4

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE			R	h_N^2	h_B^2
E_1	D_R	H_R			
141	138	284[x]	143	1.92	0.76
141	294	1006[x]	287	1.47	0.74
141	193	470[x]	191	1.72	0.72
140	382	1488[x]	377	1.34	0.70
138	3266[x]	907	214	-0.60	0.60
139	16859	5933	813	0.25	0.66
136	22481[x]	1894	430	-0.23	0.58
139	6697	16348	1482	0.38	0.63
144	67	97[x]	90	2.41	0.96
144	150	446[x]	236	1.81	0.96
144	85	147[x]	122	2.22	0.95
144	184	676[x]	343	1.68	0.95
144	375[x]	190	56	-1.80	0.93
144	∞[x]	962	246	-0.08	0.94
144	909[x]	310	85	-1.15	0.92
144	94401	1397	396	0.11	0.93

TABLE 12B.5

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE			R	h_N^2
E_1	D_R	M		
139	4306	236	154	0.74
138	1316	622	334	0.72
138	3515	358	208	0.70
136	1474	860	439	0.68
129	221	728[x]	244	0.60
133	397	27066[x]	784	0.66
128	309	1636[x]	462	0.58
132	510	∞[x]	1456	0.63
144	274	97	92	0.96
144	190	358	270	0.95
144	258	141	127	0.95
144	194	519	396	0.94
142	44	50[x]	76	0.94
144	102	899[x]	256	0.94
142	50	73[x]	117	0.93
142	107	1010[x]	429	0.93

TABLE 12B.6

POPULATION		TRUE PARAMETER VALUES			
h_N^2	ρ	E_1	E_2	D_R	D_R''
0.5	1.0	56.25	56.25	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00
0.5	1.0	56.25	56.25	225.00	112.50
0.5	0.5	56.25	56.25	225.00	112.50
0.5	-1.0	56.25	56.25	225.00	225.00
0.5	-0.5	56.25	56.25	225.00	225.00
0.5	-1.0	56.25	56.25	225.00	112.50
0.5	-0.5	56.25	56.25	225.00	112.50
0.9	1.0	11.25	11.25	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00
0.9	1.0	11.25	11.25	405.00	202.50
0.9	0.5	11.25	11.25	405.00	202.50
0.9	-1.0	11.25	11.25	405.00	405.00
0.9	-0.5	11.25	11.25	405.00	405.00
0.9	-1.0	11.25	11.25	405.00	202.50
0.9	-0.5	11.25	11.25	405.00	202.50

SAMPLE SIZE NEEDED TO DETECT: R h_N^2

D_R'	E_1	D_R	D_R''	D_R'		
225.00	116	197	1054	109	1667	0.63
112.50	128	211	947	247	1950	0.61
159.10	118	200	3865	121	1705	0.62
79.55	127	209	3899	264	1906	0.61
-225.00	128	221	1479	343[x]	1759	0.60
-112.50	132	215	1124	1937[x]	1978	0.61
-159.10	129	220	5839	536[x]	1787	0.60
- 79.55	130	215	4614	5385[x]	1937	0.61
405.00	123	84	509	97	1732	0.93
202.50	142	85	377	249	2153	0.92
286.38	135	85	1318	103	1989	0.93
143.19	142	85	1055	274	2150	0.93
-405.00	139	85	716	157[x]	1795	0.92
-202.50	142	85	476	412[x]	2153	0.92
-286.38	141	85	1783	174[x]	2016	0.92
-143.19	142	85	1278	482[x]	2249	0.92

TABLE 12B.7

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.00
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT:				R	h_N^2	h_B^2
E_1	E_2	D_R	H_R			
273	71	7315	18159[x]	285	0.17	0.07
267	314	1109	6863[x]	576	0.76	0.49
272	107	2240	5461[x]	453	0.40	0.17
262	315	1398	9635[x]	1196	0.69	0.45
129	582[x]	∞	4514	389	0.05	0.59
236	9247	99812	3578	752	0.10	0.65
164	2196[x]	∞	4756	546	0.06	0.59
233	4127	27339	6043	1535	0.20	0.61
286	38	∞	∞ [x]	138	0.01	0.00
285	1011	163	509[x]	263	1.74	0.93
285	174	107	122[x]	144	1.86	0.66
287	941	201	758[x]	427	1.61	0.92
44	51[x]	608[x]	153	138	-0.66	0.89
285	2587	40484[x]	866	246	-0.15	0.93
60	74[x]	59038[x]	762	194	-0.11	0.89
285	2635	∞	1230	406	0.05	0.92

TABLE 12B.8

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	122.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-122.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT:				R	h_N^2	h_B^2
E_1	E_2	D_R	I			
241	61	164	209^x	401	0.63	0.01
273	300	178	543^x	684	0.61	0.35
247	77	206	323^x	763	0.58	0.05
270	304	224	1044^x	1535	0.56	0.32
270	354^x	291	403	474	0.51	1.17
275	3122^x	232	1067	1431	0.55	0.95
268	687^x	324	591	858	0.49	1.06
273	9130^x	278	1600	3709	0.52	0.86
239	43	47	47^x	209	0.94	0.00
286	302	42	354^x	340	0.94	0.51
273	49	48	55^x	360	0.93	0.05
286	304	43	372^x	683	0.93	0.50
239	50^x	47	50	150	0.90	1.82
239	456^x	39	395	548	0.92	1.38
239	69^x	47	67	244	0.89	1.74
239	524^x	42	442	1201	0.91	1.35

TABLE 12B.9

POPULATION		TRUE PARMATETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT: R h_N^2

E_1	E_2	D_R	M		
270	75	∞	11993	282	0.08
266	396	1095	2973	603	0.51
272	122	21292	4111	446	0.19
261	378	1264	5330	1246	0.47
128	571[x]	956	39973	431	0.55
242	9002	339	7509[x]	719	0.65
180	3537[x]	512	12017[x]	535	0.57
241	3772	397	9176[x]	1485	0.60
286	38	∞	∞	138	0.00
285	1111	191	400	304	0.93
285	209	1258	99	146	0.72
285	1017	196	583	503	0.92
50	69[x]	1854	135	188	0.72
286	2243	98	749[x]	261	0.93
285	4849	50	73[x]	114	0.92
285	2215	103	829[x]	455	0.92

TABLE 12B.10

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	122.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZES NEEDED TO DETECT:				R	h_N^2	h_B^2
E_1	D_R	H_R	M			
141	7478[x]	5838	1145	143	-0.24	0.74
141	13072[x]	6600	1637	312	-0.10	0.72
140	12334[x]	8229	1409	196	-0.07	0.70
140	24824[x]	9668	2060	417	0.04	0.68
138	38320	4333	2069[x]	236	-0.03	0.61
139	∞ [x]	6819	42735	780	0.05	0.66
138	25399	10585	6010[x]	441	0.16	0.59
139	∞	12208	36286	1447	0.17	0.63
144	5884[x]	4336	915	84	-0.31	0.95
144	14785[x]	5358	1415	247	0.01	0.95
144	11604[x]	6520	1126	118	0.02	0.95
144	2185[x]	9204	1915	370	0.28	0.94
144	1997	895	53[x]	72	-0.17	0.94
144	1486	5343	4886[x]	251	0.36	0.94
144	978	4050	79[x]	110	0.38	0.93
144	4025	17790	3475[x]	414	0.62	0.93

TABLE 12B.11

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT: R h_N^2

E_1	E_2	D_R	D_R''	D_R'		
238	1163	395	877	197	-	0.5
275	1190	394	752	676	-	0.5
245	1160	394	2590	280	-	0.5
272	1172	393	2349	1011	-	0.5
280	943	363	1087	239x	-	0.5
282	1126	385	854	765x	-	0.5
280	972	366	3053	320x	-	0.5
281	1105	383	2578	1096x	-	0.5
239	1371	92	494	102	-	0.9
286	1418	92	364	278	-	0.9
273	1405	92	1239	111	-	0.9
286	1417	92	982	320	-	0.9
282	1042	91	687	150x	-	0.9
286	1412	77	460	373x	-	0.9
285	1253	91	1647	163x	-	0.9
286	1409	77	1186	421x	-	0.9

TABLE 12B.12

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT:

E_1	D_R	M	D_R''	D_R'	R	h_N^2
139	716	15899	1164	116	1758	0.62
138	901	∞	962	296	1869	0.61
138	734	20170	4676	134	1769	0.62
136	862	∞	4051	334	1846	0.61
128	740	14729	1515	337^x	1829	0.60
134	920	∞	1132	2244^x	1898	0.61
128	759	18479	252	531^x	1833	0.60
133	880	∞	4804	5681^x	1875	0.61
144	95	7326$_x$	517	98	2042	0.93
144	255	∞^x	379	291	2154	0.92
144	103	26655$_x$	1353	105	2043	0.93
144	224	∞^x	1067	340	2054	0.92
142	97	6623$_x$	720	454^x	2044	0.93
142	256	∞^x	479	408^x	2056	0.92
142	108	18049	1813	171^x	2045	0.93
142	229	∞	1290	573^x	2054	0.92

TABLE 12B.13

POPULATION		TRUE PARAMETER VALUES					
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'	E_1
0.5	1.0	56.25	56.25	225.00	225.00	225.00	281
0.5	0.5	56.25	56.25	225.00	225.00	112.50	281
0.5	1.0	56.25	56.25	225.00	112.50	159.10	279
0.5	0.5	56.25	56.25	225.00	112.50	79.55	286
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00	275
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50	279
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10	275
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55	284
0.9	1.0	11.25	11.25	405.00	405.00	405.00	286
0.9	0.5	11.25	11.25	405.00	405.00	202.50	286
0.9	1.0	11.25	11.25	405.00	202.50	286.38	286
0.9	0.5	11.25	11.25	405.00	202.50	143.19	286
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00	286
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50	286
0.9	-1.0	11.25	11.25	405.00	405.00	-286.38	286
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19	286

SAMPLE SIZE NEEDED TO DETECT:				R	h_N^2	h_B^2
E_2	D_R	H_R	I			
63	537	∞ x	212 x	383	0.65	0.01
383	1364	∞ x	2237 x	657	0.67	0.36
80	839	∞ x	331 x	729	0.60	0.05
377	1729	∞ x	2785 x	1472	0.61	0.33
467 x	16223 x	1977	482	528	-0.28	1.15
4923 x	15229 x	14529	1441	1481	0.27	0.93
893 x	∞ x	4723	723	907	-0.01	1.04
15737 x	9845	32799	2146	3853	0.34	0.84
43	53	∞ x	47 x	200	0.95	0.00
467	390	11320 x	555 x	333	1.12	0.55
50	77	∞ x	56 x	344	0.94	0.05
436 x	445 x	22593 x	539 x	669	1.06	0.53
55 x	165 x	108	54	143	-2.00	1.82
617 x	4179	5399	528	570	0.50	1.35
94 x	1649 x	402	73	234	-0.78	1.73
696 x	3200	7952	579	1274	0.57	1.33

TABLE

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.10
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.00
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZES NEEDED TO DETECT:					R	h_N^2	h_B^2
E_1	E_2	D_R	H_R	M			
273	85	41101^x	38409	21710	264	-0.29	0.80
267	433	37767^x	13417	5621	568	-0.06	0.52
269	128	∞ ^{x}	∞	32193	428	0.04	0.18
262	409	∞ ^{x}	49280	11706	1182	0.19	0.47
132	540^x	26114^x	3766	7144	428	-0.13	0.57
243	11530	∞	7604	∞ ^{x}	718	0.07	0.65
181	2906^x	∞	7741	∞ ^{x}	521	0.06	0.59
241	4592	1081	1600	327^x	1468	0.23	0.61
286	43	18552^x	18507	18277	128	-0.59	0.00
285	1131	15621^x	5550	1531	276	0.00	0.93
290	209	16881^x	14506	1702	134	0.02	0.68
285	1042	25443^x	10664	2191	466	0.30	0.91
62	72^x	1225^x	455	168	224	0.22	0.88
286	2313	2054	5625	3454^x	254	0.37	0.93
286	5561	899	4613	79^x	107	0.41	0.92
285	2276	3470	21638	2482^x	436	0.64	0.92

TABLE 12B.

POPULATION		TRUE PARAMETER VALUES				
h_N^2	ρ	E_1	E_2	D_R	D_R''	D_R'
0.5	1.0	56.25	56.25	225.00	225.00	225.00
0.5	0.5	56.25	56.25	225.00	225.00	112.50
0.5	1.0	56.25	56.25	225.00	112.50	159.19
0.5	0.5	56.25	56.25	225.00	112.50	79.55
0.5	-1.0	56.25	56.25	225.00	225.00	-225.00
0.5	-0.5	56.25	56.25	225.00	225.00	-112.50
0.5	-1.0	56.25	56.25	225.00	112.50	-159.10
0.5	-0.5	56.25	56.25	225.00	112.50	- 79.55
0.9	1.0	11.25	11.25	405.00	405.00	405.00
0.9	0.5	11.25	11.25	405.00	405.00	202.50
0.9	1.0	11.25	11.25	405.00	202.50	286.38
0.9	0.5	11.25	11.25	405.00	202.50	143.19
0.9	-1.0	11.25	11.25	405.00	405.00	-405.00
0.9	-0.5	11.25	11.25	405.00	405.00	-202.50
0.9	-1.0	11.25	11.25	405.00	202.50	-286.38
0.9	-0.5	11.25	11.25	405.00	202.50	-143.19

SAMPLE SIZE NEEDED TO DETECT:					R	h_N^2	h_B^2
E_1	E_2	D_R	I	M			
279	66	306	214^x	∞	383	0.63	0.01
278	534	893	3346^x	26549	657	0.60	0.38
276	85	454	337^x	∞	728	0.58	0.05
276	480	994	3506^x	∞	1471	0.56	0.34
268	639^x	18700	602	3731	508	0.45	1.39
275	2264^x	4991	1150	5749	1680	0.52	1.11
267	1168^x	6946	894	9137	876	0.46	1.20
273	6332^x	3360	1918	16680	4220	0.50	0.95
286	43	48	107^x	∞	200	0.94	0.00
286	790	208	967^x	4003	332	0.94	0.61
286	50	55	56^x	∞	344	0.93	0.05
286	591	195	738^x	12643	666	0.93	0.56
286	194^x	457	184	3020	154	0.90	2.06
286	498^x	396	454	1988	650	0.92	1.63
286	278^x	430	260	8134	242	0.89	1.89
286	784^x	376	686	6548	1335	0.91	1.50

REFERENCES

Allen, R.D., Bixler, H.H., Connor, W.L. and Graham, F.B. (1946). Metropolitan Achievement Tests, Advanced Battery - complete R.S.T.U. and V. World Book Company, Yonders on Hudson.

Anastasi, A. (1949). Differential Psychology. New York: MacMillan.

Anderson, V.E. (1974). Genetics and intelligence. In J. Wortis (Ed.) Mental Retardation (and developmental disabilities) an annual review. Vol. 6, New York: Brunnerl Mazel. 20-43.

Bartlett, M.S. (1937). Journal of the Royal Statistical Society, Supplement, 4, 137.

Block, J.B. (1968). Heredity components in the performance of twins on the Wais. In Vandenberg, S.G. (Ed.), Progress in Human Behaviour Genetics, John Hopkins Press, Baltimore.

Block, D.R. (1968). Components of Heritable Variation in mental test scores. In Vandenberg, S.G. (Ed.) Progress in Human Behaviour Genetics, John Hopkins Press, Baltimore.

Block, N.S. and Dworkin, G. (1974a). IQ, heritability and inequality (Part 1). Philosophy and Public Affairs, 3, 331-409.

Block, N.S. and Dworkin, G. (1974b). IQ, heritability and inequality (Part 2). Philosophy and Public Affairs, 4, 40-99.

Bock, R.D. and Kolakowski, D. (1973). Further evidence of sex-linked major gene influence on human spatial visualising ability. American Journal of Human Genetics, 25, 1-14.

Broadhurst, P.L., Fulker, D.W. and Wilcock, J. (1974). Behavioural genetics. Annual review of psychology, 4, 312-319.

Bronfenbrenner, U. (1967). The psychological costs of quality and equality in education. Child Development, 38, 909-925.

Bruni, P. and Eysenck, H.J. (1976). The structure of social attitudes - an Italian sample. Psychological Reports, 38, 956-958.

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

Bulmer, M.G. (1970). The Biology of Twinning in Man. London: Oxford University Press.

Burks, B.S. (1928). The relative influence of nature and nurture upon mental development: a comparative study of foster-parent child resemblance and true parent-child resemblance. In 27th Yearbook of the National Society for the Study of Education, Part I, 219-321. Bloomington: Public School Publishing.

Burt, C. (1912). The inheritance of mental characteristics. Eugenics Review, 4, 168-204.

Burt, C. (1940). The Factors of the Mind. London: University of London Press Ltd.

Burt, C. (1966). The genetic determination of differences in intelligence: a study of monozygotic twins reared together and apart. British Journal of Psychology, 57, 137-153.

Burt, C. (1972). Inheritance of general intelligence. American Psychologist, 27, 175-190.

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

Carlson, M.B. (1934). Attitudes of undergraduate students. Journal of Social Psychology, 5, 202-212.

Carroll, J.B. (1953). An analytical solution for approximating simple structure in factor analysis. Psychometrika, 18, 23-38.

Carthy, J.D. and Ebling, F.J. (Eds.) (1964). The Natural History of Aggression. London: Academic Press.

Cattell, R.B. (1957). Personality and Motivation Struction and Measurement. London: Harrap.

Cattell, R.B. (1959). Handbook for theCulture fair Intelligence Test: A Measure of 'g'. Champaign, Illinois: Institute for Personality and Ability Testing.

Cattell, R.B. (1960). The multiple abstract variance equations, and solutions for nature: nurture research on continuous variables. Psychological reviews, 67, 353-372.

Cattell, R.B. (1965a). The Scientific Analysis of Personality. London: Penguin.

Cattell, R.B. (1965b). Methodological and conceptual advances in evaluating hereditary and environmental influences and their interaction. In S.G. Vandenberg (ed.) Methods and Goals in Human Behaviour Genetics. New York: Academic Press.

Cattell, R.B. and Eber, H.W. (1967). Handbook of Sixteen Personality Factor Questionnaire. Champaign, Illinois: Institute of Personality and Ability Testing.

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

Cederlof, R., Friberg, L. Jonsson, E. and Kaij, L. (1961). Studies on similarity diagnosis in twins with the aid of mailed questionnaires. Acta Genetica et Statistica Medica (Basel), 11, 338-362.

Chein, I. (1948). Behaviour theory and the behaviour of attitudes. Psychological Reviews, 55, 175-188.

Christensen, P.R. and Guilford, J.P. (1955). Test of Logical Reasoning. Los Angeles, California: Sheridan Supply Company.

Conrad, H.S. and Jones, H.E. (1940). A second study of familial resemblance in intelligence: environmental and genetic implications of parent-child and sibling correlations in the total sample. Yearbook of the National Society for the Study of Education, 39, 97-141.

Coulter, T. (1953). An experimental and statistical study of the relationship of prejudice and certain personality variables. Unpublished. Ph.D. Thesis, University of London.

Dahlberg, G. (1926). Twin births and twins from a hereditary point of view. Stockholm: Tidens.

Dator, J. (1969). Measuring attitudes across cultures: a factor analysis of the replies of Japanese judges to Eysenck's inventory of Conservative-Progressive Ideology. In G. Schubert and D.J. Danielski (Eds.), Comparison of judicial behaviour. New York: Oxford University Press. 164-189.

Dawkins, Richard. (1976). The Selfish Gene. Oxford: Oxford University Press.

Dixon, W.J. (Editor). (1973). BMD. Biomedical Computer Programs. Los Angeles, California: University of California Press.

Doob, L.W. (1947). The behaviour of attitudes. Psychological Reviews, 54, 135-156.

Dreger, R.M. and Miller, K.S. (1960). Comparative psychological studies of Negroes and Whites in the UnitedStates. Psychological Bulletin, 57, 361-402.

Dreger, R.M. and Miller, K.S. (1968). Comparative psychological studies of Negroes and Whites in the UnitedStates: 1959-1965. Psychological Bulletin Monograph Supplement, 70 (3, pt.2).

Drillien, C.M. (1961). A longitudinal study of the growth and development of prematurely and maturely born children. Part VII: Mental development 2-5 years. Arch. Dis. Childh. 36, 233.

Du Bois, P.H. and Gleser, G. (1948). The object-aperture test; a measure involving visualisation in three dimensions. American Psychologist 3, 363.

Eaves, L.J. (1969). The genetic analysis of continuous variation: a comparison of experimental designs applicable to human data. British Journal of Mathematical and Statistical Psychology, 22, 131-147.

Eaves, L.J. (1970). The genetic analysis of continuous variation: a comparison of experimental designs applicable to human data. II. Estimation of heritability and comparison of environmental components. British Journal of Mathematical and Statistical Psychology, 23, 189-98.

Eaves, L.J. (1970). Aspects of human psychogenetics. Unpublished.Ph.D. Thesis, University of Birmingham.

Eaves, L.J. (1972). Computer simulation of sample size and experimental design in human psychogenetics. Psychological Bulletin, 77, 144-152.

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

Eaves, L.J. (1976 . A model for sibling effects in man. Heredity, 36, 205-215.

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

Eaves, L.J. (1977). Inferring the causes of human variation. Journal of the Royal Statistical Society. Series A (General), 140, 324-355.

Eaves, L.J. and Eysenck, H.J. (1974). Genetics and the development of social attitudes. Nature, 249, 288-289.

Eaves, L.J. and Eysenck, H.J. (1975). The nature of extraversion; a genetical analysis. Journal of Personality and Social Psychology, 32.

Eaves, L.J. and Eysenck, H.J. (1976). Genetical and environmental components of inconsistency and unrepeatability in twins' response to a neuroticism questionnaire. Behaviour Genetics, 6, 145-160.

Eaves, L.J. and Eysenck, H.J. (1977). A genetic model for psychoticism. Advances in Behaviour Research and Therapy, 1, 5-26.

Eaves, L.J. and Jinks, J.L. (1972). Insignificance of evidence for differences in heritability of I.Q. between races and social classes. Nature, 240, 84-88.

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

Elston, R.C. and Stewart, J. (1971). A general model for the genetic analysis of pedigree data. Human Heredity, 21, 523-542.

Erlenmeyer-Kimling, L. and Jarvik, L.F. (1963). Genetics and intelligence: a review. Science, 149, 320-321.

Eysenck, H.J. (1944). General social attitudes. Journal of Social Psychology, 19, 207-227.

Eysenck, H.J. (1947). Primary Social attitudes. I. The organisation and measurement of social attitudes. Int. J. Opin. Attit. Res., 1, 49-84.

Eysenck, H.J. (1951). Primary social attitudes as related to social class and political party. Br. J. Sociol., 2, 198-209.

Eysenck, H.J. (1952). The Scientific Study of Personality. London: Routledge and Kegan Paul.

Eysenck, H.J. (1953). Primary social attitudes: a comparison of attitude patterns in England, Germany and Sweden. Journal of Abnormal and Social Psychology, 48, 563-568.

Eysenck, H.J. (1954). The Psychology of Politics. London: Routledge and Kegan Paul.

Eysenck, H.J. (1956). The psychology of politics and the personality similarities between fascists and communists. Psychological Bulletin, 53, 431-438.

Eysenck, H.J. (1968). Genetics and personality. In Eugenics Soc. Symp. Proc. 4, 163-179.

Eysenck, H.J. (1975). The structure of social attitudes. British Journal of Social and Clinical Psychology, 14, 323-331.

Eysenck, H.J. (1976a). Sex and Personality. London: Open Books.

Eysenck, H.J. (1976b). Structure of social attitudes. Psychological Reports 39, 463-466.

Eysenck, H.J. and Eysenck, S.B.G. (1969). Personality Structure and Measurement. London: Routledge and Kegan Paul.

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

Ferguson, G.A. (1954). The concept of parsimony in factor analysis. Psychometrika, 19, 281-290.

Ferguson, L.W. (1939). Primary social attitudes. Journal of Psychology, 8, 217-223.

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

Fisher, R.A. (1921). On the mathematical foundations of theoretical statistics. Philosophical Transactions (A), 222, 309.

Fisher, R.A. (1925). Theory of Statistical Estimation. Proc. Camb. Phil. Soc., 22, 700.

French, J.W., Elkstrom, R.B. and Price, L.A. (1963). Manual for a Kit of Reference Tests for Cognitive Factors. Princeton, N.J.: Educational TestingService.

Fruchter, B. (1954). Introduction to Factor Analysis. New York: Van No Strand.

Fuller, J.L. and Thomson, W.R. (1960). Behaviour Genetics. New York: John Wiley.

Galton, F. (1869). Hereditary Genius: An enquiry into its laws and consequences. London: MacMillan.

Garrett, H.E. (1946). A Developmental Theory of Intelligence. American Psychologist, 1, 372-378.

Garrison, R.J., Anderson, V. Elving and Reed,Sheldon, C. (1968). Assortative marriage. Eugenics quarterly, 15, 113-127.

George, E.I. (1954). An experimental study of the relation between personal values, social attitudes and personality traits. Unpublished. Ph.D. Thesis, University of London.

Goodman, L.A. and Kruskal, W.H. (1954). Measure of association for cross-classifications. Journal of theAmerican Statistical Association, 49, 732-764.

Goodman, L.A. and Kruskal, W.H. (1963). Measure of association for cross-classifications III. Approximate sampling theory. Journal of the American Statistical Association, 58, 310-364.

Guilford, J.P. (1967). The Nature of Human Intelligence. New York: McGraw-Hill.

Guilford, J.P. and Lacey, J.I. (1947). Printed Classification Tests. A.A.F. Aviat. Psychol. Prog. Res. Rep. No. 5, Washington: US Government Printing Office.

Haldane, J.B.S. (1946). The interaction of nature and nurture. Am. Eugenics, 13, 197-205.

Harman, H.H. (1967). Modern Factor Analysis (second edition). Chicago and London: The University of Chicago Press.

Hartlage, L. C. (1970). Sex-linked inheritance of spatial ability. Perceptual Motor Skills, 31, 610.

Heim, A. (1970). Intelligence and Personality. Their assessment and relationship. London: Penguin.

Heim, A.W., Povey, R.N. and Watts, K.P. (1965). An attempt to measure some aspects of temperament by means of the word-in-context and self-judging vocabulary test. Journal of General Psychology, 72, 285-294.

Hendrickson, A.E. and White, P.O. (1964). Promax: a quick method for rotation to oblique simple structure. British Journal of Statistical Psychology, 17, 65-70.

Hertzka, A.F. and Guilford, J.P. (1955). Test of Logical Reasoning. Los Angeles, California: Sheridan Supply Company.

Hewitt, J. (1974). An analysis of data from a twin study of social attitudes. Unpublished. Ph.D. thesis, University of Birmingham.

Hewitt, J.K., Eysenck, H.J. and Eaves, L.J. (1977). Stgucture of social attitudes after 25 years: a replication. Psychological Reports, 40, 183-188.

Hirsch, N.D.M. (1930). Twins: heredity and environment. Cambridge: Harvard University Press.

Hope, K. (1968). Methods of Multivariate Analysis. London: University of London Press.

Huntley, R.M.C. (1966). Heritability of intelligence. In J.E. Meade and A.S. Parkes (Eds.) Genetics and environmental factors in human ability. London: Oliver and Boyd Ltd.

Husén, T. (1959). A methodological study. Psychological Twin Research 1, Stockholm: Almqvist and Wiksell.

Husén, T. (1969). Abilities of Twins. Scandinavian Journal of Psychology 1, 125-135.

Insel, P.M. (1971). Unpublished Ph.D. Thesis, University of London

Insel, P.M. (1974). Maternal effects in personality. Behaviour Genetics 4, 133-143.

Jarvik, L.F, and Erlenmeyer-Kimling, L. (1967). Survey of familial correlations in measured intellectual functions. In J. Zubin and G.A.Jervis (Eds.) Psychopathology of mental development. New York: Grune and Stratton.pp447-459.

Jencks, C. (1972). Inequality. A reassessment of the Effect of Family and Schooling in America. New York: Basic Books.

Jenkins, T.N., Coleman, J.H., Fagin, H.T. (1959). How well do you know yourself? Form NE-21. Manual of Instructions. New York: Executive Analysis Corporation.

Jensen, A.R. (1967). Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. Proceedings of the National Academy of Sciences, 58, 149-156.

Jensen, A.R. (1969). How much can we boost I.Q. and scholastic achievement? HarvardEducational Review, 39, 1-123.

Jensen, A.R. (1970). I.Q.'s of identical twins reared apart. Behaviour Genetics, 1, 133-148.

Jensen, A.R. (1972). Genetics and Education. London: Methuen.

Jensen, A.R. (1973a). Level I and Level II abilities in three ethnic groups. American Educational Research Journal, 10, 263-276.

Jensen, A.R. (1973b). Educability and Group Differences. New York: Harper and Row.

Jensen, A.R. (1973c). Educational Differences. London: Methuen.

Jinks, J.L. amd Connolly, V. (1975). Determination of the environmental sensitivity of selection lines by the selection environment. Heredity, 34, 401-406.

Jinks, J.L. and Fulker, D.W. (1970). Comparison of the Biometrical Genetical, MAVA, and Classical Approaches to the analysis of human behaviour. Psychological Bulletin, 73, 311-349.

Jinks, J.L. and Perkins, J.M. (1970). Environmental and genotype-environmental components of variability , VII. Simultaneous prediction across environments and generations. Heredity, 25, 475-480.

Joereskog, K.G. (1963). Statistical Estimation in Factor Analysis. A new technique and its foundation. Uppsala: Almqvist and Wiksell.

Joereskog, K.G. (1973). Analysis of covariance structures. In P.R. Krishmaiah, editor. Multivariate Analysis III, New York: Academic Press.

Kaiser, H.F. (1958). The varimaxcriterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200.

Kamin, L.J. (1974). The Science and Politics of I.Q. Potomac, Maryland: Lawrence Erlbaum Associates.

Kasriel, J. and Eaves, L. (1976). The Zygosity of Twins: Further Evidence on the Agreement between Diagnosis by Blood Groups and Written Questionnaires. Journal of Biosocial Sciences, 8, 263-266.

Kearsey, M.J. (1970). Experimental sizes for detecting dominance variation. Heredity, 25, 529-542.

Kempthorne, O. (1957). An introduction to genetic statistics. New York: Wiley.

Kendall, M.G. (1957). A course in Multivariate Analysis. London: Griffin.

Kendall, M.G. and Stuart, A. (1961). The Advanced Theory of Statistics. Vol. 2, Inference and Relationship. London: Charles Griffin and Company Ltd.

Klein, T.W. (1974). Heritability and genetic correlation: statistical power, population comparisons and sample size. Behaviour Genetics, 4, 171-189.

Klineberg, O. (Ed.) (1944). Characteristics of the American Negro. New York: Harper and Row.

Klineberg, O. (1963). Negro-white differences in intelligence test performance. A new look at an old problem. American Psychologist, 18, 198-203.

Lange, K., Westlake, J. and Spence, M. Anne. (1976). Extensions to pedigree analysis. III. Variance components by the scoring method. Annals of Human Genetics, London, 39, 485-491.

Lawley, D.N. (1940). The estimation of factor loadings by the method of maximum-likelihood. Proceedings of the Royal Society of Edinburgh, Series A, 60, 64-82.

Lawley, D.N. (1942). Further investigations in factor estimation. Proceedings of the Royal Society of Edinburgh, Series A, 61, 176-185.

Layzer, D. (1974). Heritability analysis of IQ scores: Science or numerology? Science, 183, 1259-1266.

Levine, J., Fishman, C. and Kagan, J. (1967). Sex of Child as Determinants of Maternal Behaviour. Paper read at a meeting of the American Ortho-Psychiatric Association, Washington, D.C.

Lewontin, R.C. (1970). Race and Intelligence. Bulletin of the Atomic Scientists 26, 2-8.

Lewontin, R.C. (1974). The analysis of variance and the analysis of causes. American Journal of Human Genetics, 26, 400-411.

Lewontin, R.C. and Cockerham, C.C. (1959). The goodness-of-fit test for detecting natural selection in random mating populations. Evolution, 13, 561-4.

Li, C.C. (1955). Population Genetics. Illinois; The University of Chicago Press.

Lindzey, G., Loehlin, J., Manosevitz, M. and Thiessen, D. (1971). Behavioural genetics. Annual Review of Psychology, 22, 39-94.

Linn, R.L. (1974). Unsquared genetic correlations. Psychological Bulletin. 81, 203-206.

Loehlin, J.C. (1965). Some methodological problems in Cattell's multiple abstract variance analysis. Psychological Reviews, 72, 156-161.

Loehlin, J.C., Lindzey, G. and Spuhler, J.N. (1975). Race differences in intelligence. pp.104-107. San Francisco: W.H.Freeman.

Loehlin, J.C. and Nichols, R.C. (1976). Heredity, Environment and Personality. A study of 850 sets of Twins. Austin and London: University of Texas Press.

Lootsma, F.A. (1972). A Survey of methods for solving constrained minimisation Problems via Unconstrained Minimisation, Numerical Methods for Non-Linear Optimisation. London: Academic Press.

Lord, F.M. and Novick, M.R. (1968). Statistical Theories of Mental Test Scores. Reading, Massachusetts: Addison-Wesley.

Lorenz, K. (1966). On Aggression. London: Harcourt, Brace and World.

MacKinnon, D.W. and Henlé, M. (1948). Experimental Studies of Psychodynamics. Cambridge: Harvard University Press.

Martin, N.G. (1976). The Classical Twin Study in Human Behaviour Genetics. Unpublished dissertation.

Martin, N.G. and Eaves, L.J. (1977). The genetical analysis of covariance structure. Heredity, 38, 79-95.

Martin, N.G. and Eysenck, H.J. (1976). Genetic factors in sexual behaviour. In Eysenck, H.J. Sex and Personality. Chapter 6, London: Open Books.

Martin, N.G. and Martin, P.G. (1975). The inheritance of scholastic abilities in a sample of twins. I. Ascertainment of the sample and diagnosis of zygosity. Annals of Human Genetics, 39, 213-218.

Mather, K. (1943). Variation and selection of polygenic characters. Journal of Genetics, 41, 159-193.

Mather, K. (1949). Biometrical Genetics (1st edn.). London: Methuen.

Mather, K. (1974). Non-allelic interaction in continuous variation of randomly breeding populations. Heredity, 32, 414-419.

Mather, K. (1975). Genotype x environmental interactions. II. Some genetical considerations. Heredity, 35, 31-53.

Mather, K. and Caligari, P.D.S. (1975). Genotype-environment interaction. III. Interactions in Drosophila melanogaster. Proc. Roy. Soc. Lond. B. 191, 387-411.

Mather, K. and Jinks, J.L. (1963). Correlations between relatives arising from sex-linked genes. Nature, 198, 314-315.

Mather, K. and Jinks, J.L. (1971). Biometrical Genetics. London: Chapman and Hall.

Mather, K. and Morley-Jones, R. (1958). Interaction of genotype and environment in continuous variation. I. Descriptions. Biometrics, 14, 343-359.

Maxwell, A.E. (1961). Recent trends in factor analysis. Journal of the Royal Statistical Society, Series A, 124, 49-59.

Maxwell, A.E. (1964). Calculating maximum-likelihood factor loadings. Journal of the Royal Statistical Society, Series A, 127, 238-241.

McNemar, ?. (1942). The revision of the Stanford-Binet Scale: an Analysis of the Standardisation Data. Boston: Houghton Mifflin.

Mehrota, S.N. and Maxwell, J. (1949). The intelligence of twins. A comparative study of eleven-year old twins. Population studies, 3, 295.

Melvin, D. (1955). An experimental and statistical study of two primary social attitudes. Unpublished. Ph.D.Thesis, University of London.

Mitra, S.K. (1958). On the limiting power function of the frequency chi-square test. Annals of Mathematical Statistics. 29, 1221-33.

Mittler, P. (1971). The Study of Twins. London: Penguin.

Mooney, C.M. (1957). Age in the development of closure ability in children. Canadian Journal of Psychology, 11, 219-226.

Moran, P.A.P. (1973). A note on heritability and the correlation between relatives. Annals of Human Genetics, 37, 217.

Morrison, D.F. (1967). Multivariate Statistical Methods. London: McGraw-Hill.

Mukherjee, B.N. (1963). Derivation of maximum-likelihood ratio tests for Guttman quasi-simplex covariance structure. Unpublished. Ph.D.Thesis, University of North Carolina.

Nance, W.E. and Corey . (1976). Genetic models for the analysis. Genetics 83, 811-826.

Neuhaus, J. and Wrigley, C. (1954). The Quartimax Method: An analytical approach to orthogonal simple structure. British Journal of Statistical Psychology, 7, 81-91.

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

Nichols, R.C. (1965). The National Merit Twin Study. In Methods and Goals in Human Behaviour Genetics. New York: Academic Press.

Nichols, R.C. and Bilbro, Jr. W.C. (1966). The diagnosis of twin zygosity. Acta Genetica Basel, 16, 265-275.

Nie, N.H., Dale, H.B. and Hull, C.H. (1970). SPSS. Statistical Package for the Social Sciences. Second Edition. New York: McGraw-Hill Book Co.

Numerical Algorithms Group. (1976). The Nag Library Manual. Nottingham.

Osborne, R.T. and Gregor, A.J. (1968). Racial differences in heritability estimates for tests of spatial ability. Perceptual Motor Skills, 27, 735-739.

Osborne, R.T., Gregor, A.J. and Miele, F. (1968). Heritability of Factor V: Verbal Comprehension. Perceptual and Motor Skills, 26, 191-202.

Osborne, R.T. and Miele, F. (1969). Racial differences in environmental influences on numerical ability as determined by heritability estimates. Perceptual Motor Skills, 28, 535-538.

Osborne, R.T. and Suddick, D.E. (1971). Blood type gene frequency and mental ability. Psychological Reports, 29, 1243-1249.

Pearson, E.S. and Hartley, H.O. (Eds.) (1972). Biometrika tables for statisticians. Volume 2. Cambridge University Press.

Pettigrew, T.F. (1964). A profile of the Negro American. Princeton, N.J.: Van Nostrand.

Plomin, R., DeFries, J.C. and Loehlin, J.C. (1977). Genotype-environment interaction and correlation in the analysis of human behaviour. Psychological Bulletin, 84, 309-322.

Price, B. (1950). Primary biases in twin studies: a review of pre-natal and natal difference producing factors in monozygotic pairs. American Journal of Human Genetics, 2, 293-352.

Rao, C.R. (1973). Linear Statistical Inference and its Applications, 2nd ed. New York: Wiley.

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

Rao, D.C., Morton, N.E. and Yee, S. (1976). Resolution of Cultural and Biological Inheritance by Path Analysis. American Journal of Human Genetics, 28, 228-242.

Record, R.G., McKeown, T. and Edwards, J.H. (1969a). The relation of measured intelligence to birth weight and duration of gestation. Annals of Human Genetics, London, 33, 71-79.

Record, R.G., McKeown, T. and Edwards, J.H. (1969b). The relation of measured intelligence to birth order and maternal age. Annals of Human Genetics. London, 33, 61-69.

Record, R.G., McKeown, T. and Edwards, J.H. (1970). An investigation of the difference in measured intelligence between twin and single births. Annals of Human Genetics, London, 34, 11-20.

Remondino, C. (1962). Recherche sur le signification du facteur numérique. (An investigation into the psychological nature of the number factor). Revuede Psychologie Applique, 12, 62-81.

Rokeach, M. (1973). The Nature of Human Values. New York: Free Press.

Sandon, F. (1957). The relative numbers and abilities of some ten-year-old twins. Journal of the Royal Statistical Society, 120, 440.

Scarr-Salapatek, S. (1971). Race, Social Class and I.Q. Science, 174, 1285-1295.

Schiff, Von F., and Verschwer, O.V. (1933). Serologische Untersuchungen an Zwillingen. Z. Morph. Anthrop. 32, 244-9.

Schoenfeldt, L.F. (1968). The hereditary components of the Project Talent two-day test battery. Measurement and Evaluation in Guidance. 1, 130-140.

Schubert, G. Human Jurisprudence. (1975). Honolulu: University press of Hawaii.

Schull, W.J. and Neel, J.V. (1965). The Effects of Inbreeding on Japanese Children. New York: Harper and Row.

Scott, J.P. and Fuller, J.F. (1965). Genetics and Social Behaviour of the Dog. Chicago: University of Chicago Press.

Seal, H.L. (1964). Multivariate Statistical Analysis for Biologists. London: Methuen.

Shields, J. (1962). Monozygotic twins brought up apart and brought up together. London: Oxford University Press.

Shuey, A.M. (1966) The Testing of Negro Intelligence. (2nd ed.). New York: Social Science Press. (1st ed., Lynchburg, Va: J.P.Bell, 1958).

Smith, I., MacFarlane, (1954). The development of a spatial test. Durham Research Review, 5, 19-33.

Smith, I., MacFarlane and Lawes, J.S. (1960). Spatial Test 3 (Newcastle Spatial Test). London: National Foundation for Educational Research in England and Wales.

Smith, S.M. and Penrose, L.S. (1955). Monozygotic and dizygotic twin diagnosis. Annals of Human Genetics, 19, 273-289.

Snedecor, G.W. and Cochran, W.G. (1967). Statistical Methods. Ames, Iowa: The Iowa State University Press.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. American Journal of Psychology, 15, 201-293.

Spence, M.A., Westlake, J. and Lange, K. (1976). Estimation of the variance components for dermal ridge count. In preparation.

Spuhler, J.N. (1967). Behaviour and mating patterns in human populations. In Spuhler, J.N. (Ed.), Genetic diversity and human behaviour, Aldine, Chicago.

Stafford, R.E. (1961). Sex differences in spatial visualisation as evidence of sex-linked inheritance. Perceptual and Motor Skills 13, 428.

Standskov and Edlen (1946). A Study of 31 Million Multiple Births.

Terman, L.M. et al (1946). Psychological Sex Differences. In Carmichael, L. (Ed.) Manual of Child Psychology. New York: Wiley.

Thurstone, L.L. (1934). The vectors of the mind. Psychological Reviews, 41, 1-32.

Thurstone, L.L. (1938). Primary mental abilities. Psychometric Monograms, 1.

Thurstone, L.L. (1945). Multiple Factor Analysis. Chicago: University of Chicago Press.

Tizard, B. (1974). IQ and Race. Nature, 247, 316.

Tizard, B. (1975). The environment and intellectual functions. In Ebling, F.J. (Ed.) Racial Variation in Man. Institute of Biology Symposium No. 22; London: The Institute of Biology.

Urbach, P. (1974a). Progress and degeneration in the 'I.Q. debate'. I. Brit. J. Philosophy of Science, 25, 99-135.

Urbach, P. (1974b). Progress and degeneration in the 'I.Q. debate'. II. Brit. J. Philosophy of Science, 25, 235-259.

Vandenberg, S.G. (Ed.) (1965). Methods and Goals in Human Behaviour Genetics. New York: Academic Press.

Vandenberg, S.G. (1966). Contributions of twin research to psychology. Psychological Bulletin, 66, 327-352.

Vandenberg, S.G. (1967). Hereditary factors in psychological variables in man with special emphasis on cognition. In Spuhler, J.N. (Ed.) Genetic Diversity and Human Behaviour, Chicago: Aldine.

Vandenberg, S.G. (1968). Progress in Human Behaviour Genetics. Recent reports on genetic syndromes, twin studies and statistical advances. Baltimore: John Hopkins University Press.

Vandenberg, S.G. (1969). A twin study of spatial ability. Multivariate Behavioural Research, 4, 273-294.

Vandenberg, S.G. (1970). A comparison of heritability estimates of U.S. Negro and White high school students. Acta Genet. Med. Gemello, 19, 280-284.

Vandenberg, S.G. (1971). What do we know today about the inheritance of intelligence and how do we know it? In R. Cancro (Ed.) Intelligence Genetic and environmental influences. New York: Grune and Stratton, pp.182-218.

Vandenberg, S.G. (1972). Assortative Mating, or Who Marries Whom? Behaviour Genetics, 2, 127-157.

Vandenberg, S.G., Stafford, R.E. and Brown, A.M. (1968). The Louisville Twin Study. In Progress in Human Behaviour Genetics. Vandenberg S.G. (Ed.) John Hopkins Press, Baltimore.

Van der Ven, A.H.G.S. (1964). Een curve analysis van de Bourdon-Wiersma Test. (A curve analysis of the Bourdon-Wiersma Test). Nederlandsch Tijdschrift voor Psychologie, 19, 300-308.

Weinberg, W. (1901). Beitrage zur Physiologie und Pathologie der Mehrlingsgeburton beim Menschen. Pflögers Arch. ges. Physiol., 88, 346-430.

Verschuer, O.V. (1932). Serologische Untersuchungen an Zwillingen. Z.Morph. Anthrop., 32, 244-249.

Weiss, L. (1975). The asymptotic distribution of the likelihood ratio in some nonstandard cases. Journal of the American Statistical Association, 70, 204-208.

Whiteman, M. (1954). The performance of schizophrenics in social concepts. Journal of Abnormal and Social Psychology, 49, 266-271.

Willerman, L. and Churchill, J.A. (1967). Intelligence and birthweight in identical twins. ChildDevelopment, 38, 623.

Willoughby, R.R. (1927). Family similarities in mental test abilities. Genetic. Psychol. Monogr. 2, 235-275.

Willoughby, R.R. (1928). Family similarities in mental test abilities. 27th Yearbook of the National Society for the Study of Education, Part 1, 55-59.

Wilson, G. (Ed.) (1973). The Psychology of Conservatism. London: Academic Press.

Wilson, G.D. and Patterson, J.R. (1968). A new measure of Conservatism. British Journal of Social and Clinical Psychology, 7, 264-90.

Winer, B.J. (1962). Statistical principles in experimental design. London: McGraw-Hill.

Wright, S. (1934). The Method of Path Coefficients. Annals of Mathematical Statistics, 5, 161-215.

Wright, S. (1921). Systems of Mating. Genetics, 6, 144-161.