Behavior Genetics, Vol. 6, No. 2, 1976

Genetic and Environmental Components of Inconsistency and Unrepeatability in Twins' Responses to a Neuroticism Questionnaire

Lindon Eaves[1] and Hans Eysenck[2]

Received 8 Aug. 1974—Final 15 Jan. 1975

Responses of twins to an 11-item neuroticism scale extracted from questionnaires administered on two widely separated occasions were subjected to a genetic analysis. The results confirm earlier findings concerning the genetic determination of neuroticism and reveal that there is a genetic component in the inconsistency of the test measured by the interaction of subjects and test items. Variation within subjects over the 2-year period between tests was due purely to environmental factors specific to individuals. When a genetic model was fitted to the raw mean squares, there was no evidence that genetic variation was other than additive and no indication of an environmental component common to members of the same family.

KEY WORDS: neuroticism; genetic; reliability; twins.

INTRODUCTION

In studies of the genetics of human behavior it is sometimes thought appropriate that estimates of genetic and environmental parameters, especially heritability estimates, should be corrected for unreliability of measurement (e.g., Cattell *et al.*, 1957). Little attention is paid, however, to unreliability as a behavioral trait in its own right and to the detailed consequences of unreliability for a genetic analysis of human behavioral traits. We have at-

This study was supported by the Colonial Research Fund and the British Medical Research Council.

[1] Department of Genetics, University of Birmingham, Birmingham, England.
[2] Institute of Psychiatry, University of London, London, England.

tempted a basic genetic analysis of components of unreliability in a scale designed for the measurement of neuroticism.

In addition to errors of measurement in the strict sense (σ^2), we distinguish two main sources of unreliability which may, under some circumstances, enter into predictions made on the basis of subjects' total scores on a test. Thus we may identify unreliability due to the interaction of subjects and individual test items ($\sigma^2_{S\times I}$), that is, the inconsistency of the test, and we may recognize unreliability due to the interaction of subjects with occasions of testing ($\sigma^2_{S\times O}$), that is, lack of repeatability of the test.

Clearly, whether or not particular components of unreliability contribute to the variation among the test scores of a random sample of subjects will depend on the generalizations we wish to make from a particular study. If we regard both occasions and items as fixed effects, the only source of unreliability in subjects' test scores will be error in the strict sense σ^2. In such a case, corrections for unreliability which involve $\sigma^2_{S\times I}$ and $\sigma^2_{S\times O}$ could be misleading. If, however, we regard occasions and/or items as random effects, then corrections for $\sigma^2_{S\times O}$ and/or $\sigma^2_{S\times I}$ will be necessary in addition to that for σ^2. In either case, we could be interested in a more detailed analysis of the interaction since it may reveal additional behavioral traits which could become the objects of further investigation. A preliminary study, such as ours, could identify at least the major areas of concern by distinguishing interactions which are predominantly genetic in origin from those which have an environmental basis. Should we wish to regard occasions and/or items as random effects, our approach gains further force because no correction of genetic parameters of the variation between subjects is possible until it has been established whether the interaction components are themselves genetic or environmental. It is sometimes supposed that unreliability contributes to environmental sources of variation between subjects and thus leads to the underestimation of heritability. This is necessarily the case only when unreliability represents error in the strict sense. In other cases, where unreliability is estimated from interaction components, "genetic" unreliability will be confounded with genetic variation between subjects' scores and "environmental" unreliability will be confounded with the environmental differences between subjects, unless an experiment is specifically designed to estimate the relevant components of the interaction in addition to the variation between subjects.

DATA

As part of a larger study concerned with the genetics of behavior, we have administered two postal personality questionnaires (the PQ and PI) to

a large number of twin volunteers. Although the questionnaires differ considerably for most of their items, we could identify 11 items related to neuroticism (N) which were formulated identically in the two questionnaires. The interval between the administration of the PQ and PI was approximately 2 years. Altogether, 441 pairs of twins completed both questionnaires. The structure of the twin sample is given in Table I. There is a marked excess of female twins, suggesting a bias in favor of female volunteers. We observe, furthermore, that the relative frequency of MZ twins is much greater than might be expected from their relative frequencies in the British population. This is a frequent finding in twin studies which rely on voluntary participation and must lead to doubt about the generality of the conclusions. We should, however, notice that such disproportionate numbers do not, by themselves, suggest that sampling is other than random for the traits under study. Zygosity was established by blood typing for some of the twins, but most were diagnosed by questionnaire.

Twins were asked:

1. Do you differ markedly in physical appearance or coloring?
2. In childhood were you frequently mistaken by people who knew you?

If consistent replies were not given, reference was made to previous questionnaires, twins' letters, and additional information in an attempt to assess zygosity. Usually such procedures have given reliable results when validated against zygosity diagnoses based on blood groups. The precise errors involved in our study are currently under investigation (Kasriel, personal communication). The 11 N items analyzed in this study are given in Table II. Since the scale is such that "yes" answers are keyed for neuroticism, a simple analysis of the individual responses which identified sources of variation due to subjects and their interaction with items and occasions will partition the total variation into that attributable to neuroticism and the various components of unreliability.

Table I. Structure of the Twin Sample: Numbers of Pairs Completing Both Questionnaires

	Monozygotic	Dizygotic	Total
Male	51	25	76
Female	202	104	306
Opposite sex	—	59	59
Total	253	188	441

Table II. Neuroticism Items Common to Both Questionnaires and Analyzed in This Study

	Item No.		
	PI	PQ	
1.	10	3	Does your mood often go up and down?
2.	14	7	Do you ever feel "just miserable" for no good reason?
3.	22	12	Do you often worry about things you should not have done or said?
4.	25	20	Are your feelings rather easily hurt?
5.	28	94	Are you sometimes bubbling over with energy and sometimes very sluggish?
6.	35	28	Are you often troubled by feelings of guilt?
7.	38	44	Would you call yourself tense or "highly strung"?
8.	58	16	Are you an irritable person?
9.	61	40	Do you worry about awful things that might happen?
10.	70	32	Would you call yourself a nervous person?
11.	78	52	Do you worry about your health?

ANALYSIS OF VARIANCE

In Table III, we present the mean squares of the analyses of variance of the five twin groups in which we recognize the hierarchical classification of subjects into pairs and individuals within pairs. The within-pairs items for the opposite-sex dizygotic twins (DZ_{os}) have been corrected for the appropriate effects due to sex (see Table IV). It is clear from Table IV that there is a significant difference between sexes for N and that sexes differ in their mean responses to the particular items. However, there is no significant interaction of sexes and occasions of testing. We give (Table V) the deviations of the item $\times$ sex means from their expectations based on the average difference between males and females over all items for opposite-sex DZ pairs. Positive deviations indicate that an item is more variable over sexes than average, negative deviations imply that an item is more consistent over sexes than might be expected. Leaving aside items 8 and 9, for which the deviations are small, there is some suggestion that the discrimination between sexes is greater for those items which relate to instability and moodiness rather than those which relate to anxiety.

We now consider in more detail the implications of the analyses of variance in Table III. We could proceed directly to fitting a genetic model to the mean squares, but some preliminary consideration will assist in deciding on an appropriate model for the data.

There are ten independent mean squares involving the triple

Table III. Mean Squares of Analyses of Variance Within Twin Groups[a]

Item	MZ_{female}		MZ_{male}		DZ_{female}		DZ_{male}		$DZ_{opposite\ sex}$	
	df	ms	df	ms	df	ms	df	ms	df	ms
Between items (I)	10	29.4103	10	6.4225	10	14.4911	10	2.1527	10	9.5270
Between pairs (P)	201	1.9752	50	1.6501	103	1.1340	24	1.0608	58	1.4417
Within pairs (W)	202	0.5811	51	0.5673	104	1.0490	25	1.2382	58	1.0314
Between occasions (O)	1	2.4979	1	0.0218	1	5.4554	1	0.1782	1	0.7458
I × P	2010	0.2937	500	0.2716	1030	0.2653	240	0.2654	580	0.2451
I × W	2020	0.1631	510	0.1791	1040	0.2158	250	0.2202	580	0.1999
I × O	10	0.5199	10	0.0767	10	0.2593	10	0.1062	10	0.2186
P × O	201	0.1935	50	0.2382	103	0.2083	24	0.2255	58	0.1642
W × O	202	0.1837	51	0.1555	104	0.1888	25	0.1691	58	0.1974
I × P × O	2010	0.1075	500	0.1051	1030	0.1058	240	0.1056	580	0.0983
I × W × O	2020	0.0939	510	0.1006	1040	0.1056	250	0.0991	580	0.0995
Subjects × occasions (pooled over all groups)					876	0.19126				
Items × subjects × occasions (pooled over all groups)					8760	0.10197				

[a] Mean squares within pairs corrected for sex difference (see Table IV).

Table IV. Analysis of Sex Differences in Opposite-Sex DZ Twins

Item	df	ms	P^a
Sexes	1	10.3606	<0.01
Sexes × occasions	1	0.0986	NS
Sexes × items	10	0.4385	<0.02
Sexes × items × occasions	10	0.1223	NS

[a] Significance levels when items are tested against the corresponding within DZ_{os} pair items on analysis in Table III.

interaction of subjects, items, and occasions. These are the mean squares I × P × O and I × W × O for each of the five groups of twins. They all appear remarkably consistent and, in fact, are so when tested for heterogeneity (χ_9^2 = 11.96, P = 0.22). This finding supports our interpretation of the triple interaction as error in the strict sense since heterogeneity would be detected if there were any genetic component of the interaction (since we would expect the within items for the DZ twins to exceed those for MZ twins) or if there were environmental effects common to members of each pair. In the latter case, we would expect there to be equal and significant components of the triple interaction between pairs, ir-

Table V. Summary of Item × Sex Interactions for Opposite Sex Dizygotic Twins

Item	Deviation[a]
1	0.0131
2	0.0936
3	0.0428
4	−0.0250
5	0.0173
6	−0.0166
7	−0.0674
8	0.0004
9	0.0046
10	−0.0250
11	−0.0376

[a] Expressed as deviation from average sex difference over all items. Positive deviations indicate a sex difference greater than expected.

respective of zygosity. The only remaining doubt is whether environmental influences specific to individuals could inflate our estimate of error. Of this we have no test with the present design, so we have pooled our ten mean squares to give a joint estimate of σ^2.

We also find that it is legitimate to combine the ten interactions of subjects and occasions since these, too, are homogeneous ($\chi_9^2 = 4.06$, $P = 0.91$). The pooled S $\times$ O interaction mean square, however, is significant when tested against our estimate of σ^2 ($\chi_{876}^2 = 1643.07$, $P < 10^{-6}$), so we must conclude that there is a real interaction of subjects and occasions. Because the magnitude of the interaction components depends neither on zygosity nor on grouping of subjects into pairs, we can interpret such interaction as the result of experiences or endogenous behavioral fluctuations which are specific to individuals irrespective of their genotype or the shared experiences of twin pairs.

For the interaction between subjects and items, and for the variation between subjects, we obtain a different result. A preliminary investigation of the S $\times$ I interactions suggests that they are all significant when tested against the pooled error but that they are not homogeneous ($\chi_9^2 = 216.39$, $P < 10^{-6}$). We see that the mean squares between MZ pairs are consistently greater than those between DZ pairs and that the reverse is true for the mean squares within pairs. This is consistent with the interaction having at least some genetic basis. A similar pattern emerges for the variation between subjects (i.e., for "neuroticism"), but this can be interpreted without reference to the inconsistency of the test only if we are prepared to regard test items as fixed effects.

A simplified statistical model for the mean squares of a typical analysis is given in Table VI. We have assumed throughout that both subjects and occasions represent random effects, but we have indicated the expectations on both random and fixed models with respect to items. In Table VII, we provide the estimated components of variance of the individual responses calculated on the basis of both models for the five groups of twins. Bearing in mind the large errors inevitably associated with estimated components of variance, the estimates which are expected to be similar are quite consistent.

GENETIC ANALYSIS

The estimates in Table VII still do not represent the most parsimonious summary of the data. We may reparametrize our expectations of mean squares for the five analyses of variance in terms of a simple genetic model which makes explicit certain theoretical relationships which may exist between the components of variance of different analyses if the model

Table VI. Expectations of Mean Squares on Statistical Model[a]

Items	σ^2		$+ 2n\sigma^2_{IxO}$	$+ 2\sigma^2_{IxW}$	$+ 4\sigma^2_{IxP}$				$+ 4n\sigma^2_{I}$
Pairs	σ^2	$+ 11\sigma^2_{SxO}$		$+ \mathbf{2}\sigma^2_{IxW}$	$+ \mathbf{4}\sigma^2_{IxP}$	$+ 22\sigma^2_{W}$	$+ 44\sigma^2_{P}$		
Within pairs	σ^2	$+ 11\sigma^2_{SxO}$		$+ \mathbf{2}\sigma^2_{IxW}$		$+ 22\sigma^2_{W}$			
Occasions	σ^2	$+ 11\sigma^2_{SxO}$	$+ 2n\sigma^2_{IxO}$						$+ 22n\sigma^2_{O}$
I × P	σ^2			$+ 2\sigma^2_{IxW}$	$+ 4\sigma^2_{IxP}$				
I × W	σ^2			$+ 2\sigma^2_{IxW}$					
I × O	σ^2		$+ 2n\sigma^2_{IxO}$						
Pooled subjects × O	σ^2	$+ 11\sigma^2_{SxO}$							
Pooled error (S × I × O)	σ^2								

[a] Components in bold type appear only if items are not considered fixed.

Table VII. Neuroticism: Components of Variation of Individual Responses to 11 Items on Two Occasions

	Item	MZ_f	MZ_m	DZ_f	DZ_m	DZ_{os}
Pooled	σ^2			0.10197		
	σ^2_{SxO}			0.00812		
Items fixed or random	σ^2_{IxO}	0.00103	—	0.00076	0.00008	0.00099
	σ^2_{IxW}	0.03057	0.03857	0.05692	0.05912	0.04897
	σ^2_{IxP}	0.03265	0.02313	0.01238	0.01130	0.01130
Items fixed	σ^2_O	0.00052	—	0.00230	—	0.00043
	σ^2_W	0.01772	0.01709	0.03899	0.04759	0.03819
	σ^2_P	0.03168	0.02461	0.00193	—	0.00933
	k^2_I	0.03552	0.03015	0.03382	0.01883	0.03884
Items random	σ^2_O	0.00043	—	0.00223	—	0.00034
	σ^2_W	0.01494	0.01359	0.03381	0.04221	0.03374
	σ^2_P	0.02872	0.02169	0.00081	—	0.00830
	σ^2_I	0.03552	0.03015	0.03382	0.01883	0.03884

is appropriate. In this case, we specify genetic parameters for variation in neuroticism (D_{RN}) and for the interaction of subjects and items ($D_{RS\times I}$), and we specify environmental parameters for variation in neuroticism (E_{1N}), for the interaction of subjects and items ($E_{1S\times I}$), for the interaction of subjects and occasions ($E_{1S\times O}$), and for error (σ^2). In Table VIII we give expectations for the components of variance in terms of our simple model and in Table IX the expectations for the relevant mean squares of the analyses of variance in terms of our genetic and environmental components. The assumptions which are made in writing this model are given in detail elsewhere (Eaves and Eysenck, 1975), but the principal assumptions are that mating is random for the traits under study, that all genetic variation is additive, that any environmental influences are specific to individuals rather than common to twin pairs, and that there are no effects of sex linkage or sex limitation.

The model we are fitting, therefore, is the simplest possible model for the combined action of genetic and environmental influences, and one of the very few models which can be tested at all with data on twins reared together. Any complexity which is apparent rather than real lies not in the model but in the estimation procedure which we must employ to assess its validity. Studies which do not specify the precise model on which an interpretation is based and which do not, however inadequately, attempt to test the assumptions statistically may lead either to the unjustified adoption of a oversimplified model or to the equally unjustified estimation of nonsignificant parameters in an unduly complicated model.

Provided that observed and expected mean squares are in close agreement, it is a simple matter to obtain weighted least-squares estimates of our parameters which approximate to maximum likelihood estimates given that the mean squares are normally distributed. Eaves and Eysenck (1975) consider a similar application of the method in greater detail.

There are two main reasons for adopting a weighted least-squares procedure. First, it means that statistics which are based on relatively few observations (in this case, for example, the male DZ statistics) play a comparatively small part in determining the final solution. Second, it provides a statistical basis for deciding whether discrepancies in the data should be taken seriously or regarded merely as the result of sampling variation. Examination of the raw mean squares (Table III) suggests that while the data are generally consistent over sexes and for twins of both types there are anomalies such as the negative intraclass correlation for DZ males. We need some objective criterion for deciding whether such a result, based as it is on a small sample, gives us reason to doubt the general validity of the model. One criterion for such decisions is the χ^2 test of goodness of fit of

Table VIII. Genetic and Environmental Model for Variance Components

Component	Parameter					
	D_{RN}	E_{1N}	$D_{RS\times I}$	$E_{1S\times I}$	$E_{1S\times O}$	E_1
σ^2	.	.	.	.	.	1
$\sigma^2_{S\times O}$	.	.	.	.	1	.
$\sigma^2_{I\times MZW}$	.	.	.	1	.	.
$\sigma^2_{I\times MZP}$	.	.	$\frac{1}{2}$	.	.	.
σ^2_{MZW}	.	1	.	.	.	.
σ^2_{MZP}	$\frac{1}{2}$	.	.	.	.	.
$\sigma^2_{I\times DZW}$	.	.	$\frac{1}{4}$	1	.	.
$\sigma^2_{I\times DZP}$	.	.	$\frac{1}{4}$	.	.	.
σ^2_{DZW}	$\frac{1}{4}$	1	.	.	.	.
σ^2_{DZP}	$\frac{1}{4}$	.	.	.	.	.

the model which is obtained at the conclusion of the model-fitting procedure. Only if our model fails at some predetermined level are we likely to be correct in designing future research around apparent anomalies in the data. The χ_{16}^2 for testing the goodness of fit of the model was 19.03 ($P = 0.27$), indicating that our simple model gives quite an economical account of the variation in individual responses to the questionnaire. The estimates of the parameters and their standard errors are given in Table X. In Table XI, we summarize the contributions of the different sources to the variation in individual responses.

From the appropriate components of the fixed- or random-item models, we may estimate any desired reliability coefficient and obtain values of the heritability of the trait and its inconsistency. The "true" heritability of neuroticism is thus

$$\hat{h}^2 = \tfrac{1}{2}\hat{D}_{RN}/(\tfrac{1}{2}\hat{D}_{RN} + E_{1N})$$

which yields a value of 0.57 for a fixed set of items and 0.59 if we regard our items as a random selection. It is important to stress that our heritability estimate is based on components estimated by weighted least squares and that we may calculate the heritability in this way only because we have shown our simple genetic model to be adequate. Estimates obtained from correlation coefficients usually imply similar assumptions with no test of the genetic model and involve a relatively inefficient use of data.

Table IX. Expectations of Relevant Mean Squares on Simple Genetic and Environmental Model[a]

Mean square	Observed	df	Information	E_1	$E_{1S\times O}$	$E_{1S\times I}$	$D_{RS\times I}$	E_{1N}	D_{RN}
MZP_f	1.9752	201	25.7599	1	11	**2**	**2**	22	22
MZP_m	1.6501	50	9.1816	1	11	**2**	**2**	22	22
DZP_f	1.1340	103	40.0480	1	11	**2**	**1½**	22	16½
DZP_m	1.0608	24	10.6639	1	11	**2**	**1½**	22	16½
DZP_{os}	1.4417	58	13.9524	1	11	**2**	**1½**	22	16½
MZW_f	0.5811	202	299.1022	1	11	**2**	.	22	.
MZW_m	0.5673	51	79.2346	1	11	**2**	.	22	.
DZW_f	1.0490	104	47.2555	1	11	**2**	**½**	22	5½
DZW_m	1.2382	225	8.1532	1	11	**2**	**½**	22	5½
DZW_{os}	1.0314	58	27.2611	1	11	**2**	**½**	22	5½
$I \times MZP_f$	0.2937	2010	11650.8650	1	.	2	2	.	.
$I \times MZP_m$	0.2716	500	3389.0695	1	.	2	2	.	.
$I \times DZP_f$	0.2653	1030	7316.9945	1	.	2	1½	.	.
$I \times DZP_m$	0.2654	240	1703.6462	1	.	2	1½	.	.
$I \times DZP_{os}$	0.2451	580	4827.3788	1	.	2	1½	.	.
$I \times MZW_f$	0.1631	2020	37967.6268	1	.	2	.	.	.
$I \times MZW_m$	0.1791	510	7949.6683	1	.	2	.	.	.
$I \times DZW_f$	0.2158	1040	11166.0730	1	.	2	½	.	.
$I \times DZW_m$	0.2202	250	2577.9553	1	.	2	½	.	.
$I \times DZW_{os}$	0.1999	580	7257.2554	1	.	2	½	.	.
$S \times O$	0.19126	876	11973.6293	1	11	.	.	.	.
Error	0.10197	8760	421239.6777	1	.	.	.	.	.

[a] Coefficients in bold type do not apply if items are regarded as fixed.

Table X. Estimates of Genetic and Environmental Components

Parameter	$\hat{\theta}$	σ_θ
D_{RN}, items fixed	0.04943	0.00601
D_{RN}, items random	0.04391	0.00602
E_{1N}, items fixed	0.01857	0.00224
E_{1N}, items random	0.01550	0.00226
$D_{RS} \times I$	0.06075	0.00418
$E_{1S} \times I$	0.03375	0.00223
$E_{1S} \times O$	0.00812	0.00084
E_1	0.10197	0.00154

On either model for the items, we may estimate the heritability of the test inconsistency from

$$\hat{h}^2 = \tfrac{1}{2}\hat{D}_{RS\times I}/(\tfrac{1}{2}\hat{D}_{RS\times I} + E_{1S\times I})$$
$$= 0.47$$

DISCUSSION

The fact that D_{RN} and $D_{RS\times I}$ are significant indicates that there is significant genetic variation both for neuroticism itself and for the interaction of subjects and items. The former conclusion confirms a repeated finding (e.g., Shields, 1962; Jinks and Fulker, 1970; Eaves, 1973). The latter con-

Table XI. Percentage of Total Variation in Individual Responses Attributed to Random Sources

	"Neuroticism"	Inconsistency	Unrepeatability	Error	Total
		Items fixed			
Genetic ($\frac{1}{2}D_R$)	11.36	13.97	—	—	25.33
Environmental (E_1)	8.54	15.52	3.73	46.88	74.67
Total	19.90	29.49	3.73	46.88	
h^2	0.57	0.47	—	—	
		Items random			
Genetic ($\frac{1}{2}D_R$)	10.37	14.35	—	—	24.72
Environmental (E_1)	7.32	15.94	3.84	48.17	75.27
Total	17.69	30.29	3.84	48.17	
h^2	0.59	0.47	—	—	

clusion is new in the context of personality measurement and suggests that a strictly unitary model for the genetics of neuroticism measures may not be appropriate. In our case, we find that the particular pattern of "symptoms" revealed by responses to a questionnaire is itself partly genetically determined, as well as the overall predisposition to neuroticism. For other aspects of behavior, we might expect a different picture. For social attitudes, for example, in which some genetic determination of the factors is indicated (Eaves and Eysenck, 1974), we would not be surprised if the inconsistency of subjects' responses reflected cultural and specific environmental influences. Such evidence as we have, however, suggests that the inconsistency of social attitude responses also has a genetic component (Hewitt, 1974, personal communication).

With regard to the causes of change in N with time, our analysis has been restricted to directional changes only, such as may be due to aging, for example, or, in another context, to the responses of different genotypes to short-term cultural changes. We have shown that there is little evidence of any genetic component of such changes. We may also examine the genetic and environmental basis of nondirected change by considering the absolute differences between scores on two occasions, irrespective of sign. There is no particular reason why the causes of directed change, be they genetic or environmental, should be the same as those responsible for nondirected fluctuations in behavior. To examine the causes of such fluctuations, we obtained the absolute difference between the N scores of each individual on the two occasions and conducted an analysis of variance of these scores as a prelude to analysis in genetic and environmental terms. Such an analysis has to be received with some skepticism because the distribution of the derived scores is far from normal.

In Table XII, we present the mean squares and intraclass correlations for the five groups of twins. The total variances of the groups are homogeneous ($\chi_4^2 = 5.47$).

A mere inspection of the correlations suggests that no simple model is likely to fit the change scores since, apparently, two of the correlations are significantly negative. We must proceed tentatively because of nonnormality, but the correlations are nearly heterogeneous at the 5% level ($\chi_4^2 = 9.32$) and the best-fitting pooled correlation is -0.09, which again approaches significance at the 5% level. It is therefore tempting to accept the simplest conclusion that there is little evidence that anything other than individuals' specific environmental experiences contributes to nondirected changes in neuroticism. If such an explanation is thought to be unsatisfactory, then we must seek an explanation in terms of competition within families such that extreme variability of one individual evokes extreme

Table XII. Analyses of Variance of Absolute Change Scores

	MZ_{female}		MZ_{male}		DZ_{female}		DZ_{male}		DZ_{os}	
Item	ms	df	ms	df	ms	df	ms	df	ms	df
Between pairs	1.66	201	2.22	50	2.19	102	1.06	24	1.92	58
Within pairs	2.23	202	1.66	51	2.71	103	3.00	25	1.59	59
Total	1.95	403	1.94	101	2.45	205	2.04	49	1.75	117
$r_{intraclass}$	−0.147		0.144		−0.106		−0.478		0.094	

stability from another, or *vice versa*. It would in our view be unwise to take such an explanation too seriously on the basis of these data.

Our finding that test inconsistency may be under genetic control is by no means unusual in quantitative genetics. It is regularly seen (e.g., Mather, 1953; Jinks and Mather, 1955; Paxman, 1956) that variation between, for example, measurements of replicated structures in the same organism is partly under genetic control and may therefore be subject as much to the influence of natural selection as any other genetically determined trait. Attempts to correct heritability estimates for unreliability of measurement which use an inappropriate model for unreliability may, at least in principle, lead to overestimation of the genetic component of variation among subjects' true scores. Whatever the source of such interaction, the usual correction for unreliability will overestimate the contribution of error to the variation of subjects' scores on a fixed set of items. If, in addition, the inconsistency has a genetic component, it will be wrong to assume that inconsistency in the responses of subjects to a random collection of items contributes only to our estimate of environmental variation. Similar considerations may apply to lack of test repeatability for traits in which developmental factors operate between occasions of testing. Wilson (1972) has illustrated the genetic control of developmental profiles for one behavioral trait. Misleadingly high heritability estimates could result from inappropriate corrections for unreliability.

There is, therefore, no substitute for an appropriate complete experimental design and genetic analysis if the variation in test responses is to be assigned unambiguously to the appropriate genetic and environmental sources.

Given measurements which were continuous at the item level, rather than dichotomous, we would have examined in greater detail the genetic and environmental contributions to individual components of the

interaction. Such an attempt is ill-advised with our data, however, unless alternative analyses are employed.

Our analysis has allowed us to test, albeit with little power in certain instances, some of the assumptions which are usually untested in the more traditional analyses of correlations. In particular, we have tested assumptions about the additivity of gene action, the randomness of mating, and the absence of common environmental influences, sex linkage, and sex limitation (Eaves and Eysenck, 1975). Furthermore, we have tested assumptions which are usually implicit in any attempt to estimate genetic and environmental components of true scores from subjects' scores on questionnaires.

It is especially difficult to provide a powerful test of the assumption of additive gene action (Eaves, 1972), but our finding that gene action is mainly additive suggests as a working hypothesis that neuroticism may display the kind of genetic architecture associated with an evolutionary history of stabilizing selection; that is, individuals of intermediate phenotype are at a selective advantage relative to those at either extreme. Twin studies such as ours, however, provide little more than a starting point for a more detailed genetic analysis of this kind (Jinks and Fulker, 1970).

ACKNOWLEDGMENTS

We thank Mrs. J. Karriel for collection of the data and Professor J. L. Jinks for helpful suggestions on the manuscript.

REFERENCES

Cattell, R. B., Stice, G. F., and Kristy, N. (1957). A first approximation to nature–nurture ratios for eleven primary personality factors in objective tests. *J. Abnorm. Soc. Psychol.* **54**:143–159.

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

Eaves, L. J. (1973). The structure of genotypic and environmental covariation for personality measurements: An analysis of the PEN. *Brit. J. Soc. Clin. Psychol.* **12**:275–282.

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. *J. Personal. Soc. Psychol.* (in press).

Jinks, J. L., and Fulker, D. W. (1970). A comparison of the biometrical genetical MAVA and classical approaches to the analysis of human behavior. *Psychol. Bull.* **73**:311–349.

Jinks, J. L., and Mather, K. (1955). Stability in development of heterozygotes and homozygotes. *Proc. Roy. Soc. Lond. Ser. B* **143**:561–578.

Mather, K. (1953). Genetical control of stability in development. *Heredity* **7**:297–336.

Paxman, G. J. (1956). Differentiation and stability in the development of *Nicotiana rustica*. *Ann. Bot.* **20**:331–347.

Shields, J. (1962). *Monozygotic Twins,* Oxford University Press, London.

Wilson, R. S. (1972). Twins: Early mental development. *Science* **175**:914–917.