Br. J. soc. clin. Psychol. (1973), 12, pp. 275–282
With 2 text-figures
Printed in Great Britain

The Structure of Genotypic and Environmental Covariation for Personality Measurements: An Analysis of the PEN

By L. J. EAVES

Department of Genetics, University of Birmingham

Phenotypic covariation for personality measurements reflects covariation due to both genetical and environmental influences. An attempt was made to separate covariation due to genetical causes from that due to environmental causes among the responses of 101 pairs of monozygotic twins to an 80-item questionnaire providing scores of psychoticism, extraversion and neuroticism. At least part of the covariation of items known to load for extraversion could be explained by the unitary action of environmental influences. The same conclusion was reached with respect to neuroticism, but there was little evidence that environmental differences resulted in any covariation between the psychoticism items.

Although genetical influences were more important in determining the relationships among the P items there was no clear indication that psychoticism was a unitary genotypic factor. The covariation of the N items, however, could reasonably be attributed to unitary genetical influences on neurotic behaviour. It was found that more than one genetical factor was probably involved in the determination of covariation among the E items, suggesting that the apparently unitary nature of extraversion at the phenotypic level could be due to environmental rather than to genetical influences.

The most widespread approach to the genetics of human behaviour involves the genetical analysis of psychological measurements derived by a consideration of phenotypic variation. It has been recognized that behavioural traits should, as far as possible, represent unitary structures, and factor analysis has been adopted as an appropriate method of defining unitary traits (e.g. Eysenck & Prell, 1951). Defining behavioural factors on the basis of phenotypic variation alone, however, assumes that genotypic and environmental influences cooperate in producing the observed covariance structure. It is equally possible that the structures of genotypic and environmental covariation may differ and could be analysed separately.

There have been various attempts to resolve the genotypic and environmental factors of covariation for cognitive measurements (Cattell, 1963; Loehlin & Vandenberg, 1968; Bock & Vandenberg, 1968; Roudabush, 1968) and Loehlin (1965) reports one approach to the separation of genotypic and environmental factors of personality. This paper summarizes an attempt to achieve a similar resolution for personality factors defined by responses to the PEN questionnaire (H. J. Eysenck & S. B. G. Eysenck, 1968; S. B. G. Eysenck & H. J. Eysenck, 1968, 1969).

METHOD

The data on which the analysis was based were kindly supplied by Dr J. S. Price, and consisted of answers by 101 pairs of monozygotic twins to the PEN. The twins were originally ascertained by their participation in a David Frost television programme and their monozygosity was initially established with reasonable confidence by their responses to a question concerning their childhood similarity (Cederlöf *et al.*, 1961). Later, more rigorous zygosity determination for 55 of the pairs confirmed that there had been no misclassification for this part of the sample (Price, personal communication).

The data thus comprised 80 yes/no responses by 202 individuals. For each twin pair the individual responses (scored 1 or 0) were added for each item to give an estimate of the joint response of the pair. Half the difference between the responses of the individuals of a pair to an item was taken as an estimate of the within-pair difference in response to the item. Covariances were calculated between every possible pair of items for sums and differences separately, giving the between-pairs covariance matrix, **B**, and the within-pairs matrix, **W**. Since the responses are dichotomous, and not normally distributed, the tests of significance applied in the subsequent analysis cannot be interpreted very rigorously. It was not possible, in this case, to consider the scores for males and females separately, because the analysis is only applicable when the number of pairs exceeds the number of variables in the study. Data on males and females was therefore pooled to secure covariance matrices of full rank.

The within-pairs covariance matrix reflects the structure of the within-family environmental (E_1) influences on response to the items, and any interaction of such effects with the genotypic differences (GE). Jinks & Fulker (1970) discuss such biometrical considerations in greater detail for the case of single variables. The twins in this study had been separated for varying lengths of time, and any environmental differences arising from such a separation (E_2) will be confounded with the E_1 effects in the within-pairs variation. No evidence was found for this sample of any relationship between degree of separation and intra-pair difference for factor scores suggesting that E_2 effects could be discounted as components of within-pairs variation for these data (Eaves, 1970).

The between-pairs covariance matrix reflects the structure of genotypic influences on behaviour (G), but confounded with these will be any common environmental (E_2) influences which operated before the twins were separated and could not be detected with these data. Jinks & Fulker (1970) have shown by their reanalysis of Shields' twin data (Shields, 1962) that there are no E_2 effects for E and N scales similar to those of the PEN. Differences due to sex and age are also confounded with the between-pairs variation in this sample. These differences may inflate the between-pairs variance if the factor scores vary with sex and age, and may confuse the covariance structure if the factor loadings vary between sexes and with time. Sex and age differences are known for E and N (S. B. G. Eysenck & H. J. Eysenck, 1969). Sex differences were confirmed for the N factor scores from this sample but not for E. Eysenck (personal communication) has observed that females endorse the P item, number 26 ('Would it upset you to see a child or animal suffer?') more frequently than males. The P loading of this item also differs greatly between the sexes (Eysenck & Eysenck, 1968). Any subsequent interpretation of the 'genotypic' factors must be made with regard to these possible effects.

The structure of the environmental variation was investigated by extraction of the principal components of **W**. Components analysis, rather than factor analysis, was used for simplicity. This, in any case, is the initial procedure frequently adopted by the authors of the scales (Eysenck & Eysenck, 1969). The relative importance of the components assessed by Bartlett's chi-square criterion (Bartlett, 1950) and the component loadings were obtained by appropriate rescaling of the normalized eigenvectors of **W**.

A model of the genotypic covariation was provided by a canonical analysis using **B** as the hypothesis matrix and **W** as the error matrix. This analysis produces linear combinations of the observed responses which maximize the variation between pairs relative to the variation within pairs. Maxwell (1961) illustrates the application of canonical analysis to dichotomous variables. The roots of $|\mathbf{B}-\phi\mathbf{W}|=0$ were extracted and their significance tested by chi-square (see Hope, 1968).

Each latent root is the between-pairs variance for the corresponding canonical variate when the within-pair variance is standardized to unity. In terms of the genotypic and environmental components of variance, the variance between-pairs, ϕ, is $2G+E_1$, and that within-pairs is E_1, provided E_2 makes a relatively small contribution to the total variance. Age and sex differences which are under genetical control contribute to G, together with any environmental or cultural influences which are inseparably associated with the genetical variation. For a canonical variate, E_1 is standardized to unity so the between-pairs variance is $2G+1$ and the within-pairs variance is unity. The proportion of the total variance which can be ascribed to genetical causes is thus

$$h_i^2=(\phi_i-1)/(\phi_i+1)$$

for the ith canonical variate. This is the 'broad heritability' (Mather & Jinks, 1971) and is applicable to the general population assuming the sample is representative with respect to

the distribution of genotypes, environmental influences, sex and age. Common environmental influences operating before separation would inflate this estimate of heritability. Both additive and dominance variation are confounded in G (Jinks & Fulker, 1970), so it is impossible to estimate the components of gene action with these data.

A canonical vector, $\boldsymbol{\alpha}_i$, was obtained for each significant root ϕ_i, by solution of the set of simultaneous equations $(\mathbf{B}-\phi_i\mathbf{W})\boldsymbol{\alpha}_i=0$. Each vector is a set of regression weights which can be applied to the observed responses to generate scores which show the largest possible heritability. To avoid the difficulty usually experienced in the interpretation of canonical vectors they were used to transform the raw pair means and 'genotypic factor loadings' were computed directly by correlating raw and transformed variates. When all canonical variates are considered, these loadings reproduce the matrix of correlations between pair means. Ten of the genotypic factors were subsequently rotated by Varimax and Promax in an attempt to define simple structure. Three second-order factors were extracted from the oblique factors in a final attempt to clarify the genetical structure of covariation.

Two scoring scales for the three factors were supplied (Eysenck, personal communication) and the interpretation of factor structure is based on the weighting assigned to the items in the generation of factor scores.

RESULTS

All the components of $\mathbf{W}$ were equal in size within the limits of sampling error ($\chi^2_{(3320)}=1477{\cdot}63$, $P>0{\cdot}05$). This finding is consistent with the fact that the within-pair variances for the three sets of factor scores are virtually equal (Eaves, 1970). The loadings of representative items on the first three principal components are given in Table 1.* These approximate closely to the simple structure solution obtained by attempts at both orthogonal and oblique rotation. Examination of the loadings suggests that environmental factors of psychoticism, extraversion and neuroticism are not clearly defined in this sample. Only items which are sufficiently well established for inclusion in scoring scales are considered when interpreting the components. The established items are labelled in the second column of Table 1. The P items were found to have very few high loadings except on the numerically smaller components. Individual P items tended to be highly specific to individual small components, which probably reflects small correlation between the P items either because of low item reliability or low frequency of endorsement.

In order to clarify the structure of the environmental variation further, the projections of the item vectors on the plane of the first two components are represented in Fig. 1. E and N items are clearly separated in these two dimensions, suggesting that the environment may indeed operate in a unitary fashion for each of these two factors. The fact that the items are widely scattered, however, accounts for the difficulty experienced in interpreting the environmental components and may simply be attributable to the small sample size. The fragmentary nature of the P factor, however, does not support the hypothesis of a unitary factor with respect to environmental variation for psychoticism as measured by this questionnaire. A slight tendency for the P items to be spread in the direction of the N items is consistent with the correlation between P and N reported by the authors of the scales (H. J. Eysenck & S. B. G. Eysenck, 1968). This suggests that environmental influences which contribute to neuroticism might also result in variation for psychoticism, or vice versa. Further, if there are genes affecting the organisms' responsiveness to the environment, they may display a joint influence on the determination of

* Copies of full tables of the factor loadings may be obtained from the author, University of Birmingham, P.O. Box 363, Birmingham B15 2TT.

Table 1. *Loadings of selected items on the first three orthogonal factors of environmental covariation and on the first three oblique factors of genotypic covariation*

(Decimal points omitted.)

Item	Scale	Environmental			Genotypic		
		I	II	III	I	II	III
2	N	23	23	44	11	−23	−19
35	N	34	31	−13	−02	−30	−16
10	N	24	19	−00	02	−10	−36
14	N	34	18	−22	02	06	−45
28	N	01	40	−11	03	12	−33
31	N	26	29	−16	05	−03	−34
38	N	57	−01	23	04	−04	−25
58	N	46	19	−03	01	09	−34
61	N	30	30	12	11	−13	−25
73	N	39	22	−03	−11	−08	−32
50	N	−01	12	38	−06	10	−06
60	N	14	20	25	06	−05	04
76	N	54	−05	−05	00	−04	−14
78	N	30	19	08	05	03	08
45	N(−P)	10	14	11	−01	−02	−51
4	E	−44	11	−32	01	01	05
16	E	−26	35	−03	11	−02	−02
27	E	−40	30	−05	−02	00	−26
36	E	−46	08	13	07	01	−00
40	E	−51	26	02	10	−05	−09
44	−E	32	−20	−42	−03	10	−00
75	−E	62	−03	−24	08	−10	−06
1	E(N)	05	35	−03	−03	−19	−15
20	E	−07	55	−14	02	−14	04
56	−E	30	−08	11	12	−27	−04
66	E	−20	28	26	−03	−38	07
69	E	−36	39	00	−11	−23	−19
72	E	−27	15	09	00	28	−16
77	E	−48	21	31	04	−28	01
26	−P	14	02	12	80	04	00
39	−P	04	13	01	65	−13	08
19	−P	−04	−21	−17	−01	14	−19
32	P	05	08	−03	07	02	35
55	−P	−00	−19	03	−03	−09	−35
71	−P	02	01	09	−10	06	−47
47	−P	−12	−18	−09	01	26	−06
3	P	04	−05	−14	−16	−14	12
5	P	−09	−04	−20	00	−00	−03
19	P	08	08	15	17	10	−09
21	P	−01	01	07	−04	10	10
37	P	−01	07	08	07	01	−00

scores for P and N. Either of these interpretations is consistent with this analysis of **W**.

Of the 80 roots extracted in the canonical analysis the first 21 were regarded as statistically significant as the remainder were homogeneous ($\chi^2_{(4681)} = 4749{\cdot}08$; $P > 0{\cdot}05$). The significant roots accounted for 88 per cent of the trace of ($\mathbf{BW}^{-1}$) and are presented in Table 2 with their heritabilities estimated as described above. The heritabilities of these genetical factors are all extremely high.

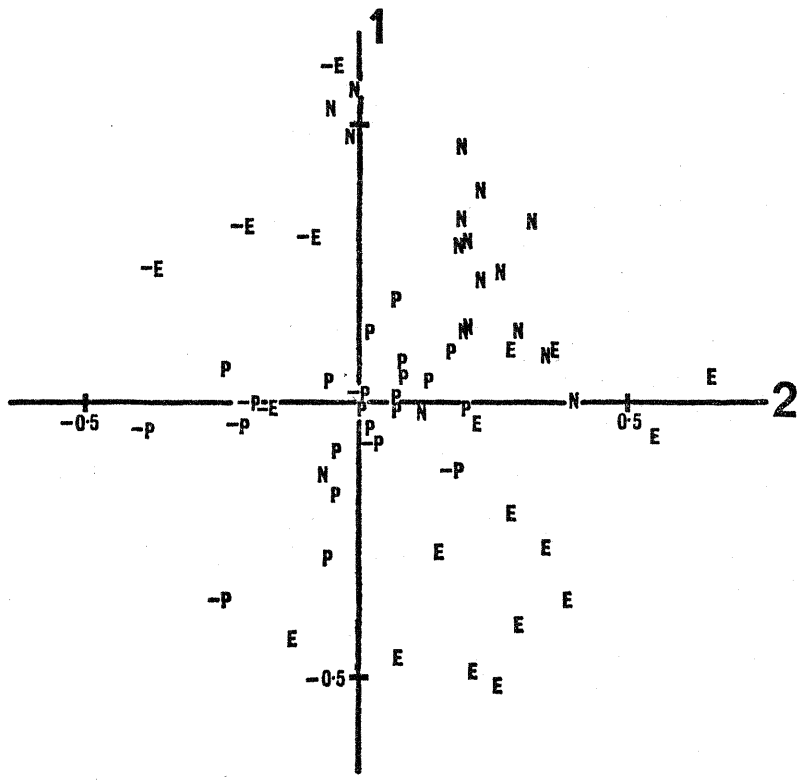

Fig. 1. The first two dimensions of environmental variation for the PEN. The position of the end of a vector is indicated by the initial letter of the scale for which an item is scored.

Table 2. *The first 21 canonical roots of the PEN data*

No.	Root (ϕ)	Proportion of trace of $\mathbf{BW}^{-1}$	'Heritability'
1	141·66	0·18	0·98
2	120·87	0·15	0·98
3	67·98	0·08	0·97
4	48·51	0·05	0·96
5	39·67	0·05	0·95
6	37·33	0·05	0·95
7	32·81	0·04	0·94
8	26·64	0·04	0·93
9	25·72	0·03	0·92
10	22·75	0·03	0·91
11	19·62	0·02	0·90
12	18·76	0·02	0·90
13	18·28	0·02	0·90
14	15·36	0·02	0·88
15	13·96	0·02	0·87
16	12·43	0·01	0·85
17	12·19	0·01	0·85
18	11·02	0·01	0·84
19	9·77	0·01	0·81
20	8·27	0·01	0·78
21	6·98	0·01	0·75

The first ten canonical variates which accounted for 73 per cent of the trace of ($\mathbf{BW}^{-1}$) were transformed to genotypic factor loadings. Table 1 gives the loading of representative items on the first three Promax factors obtained by rotation of the ten orthogonal genotypic factors. The high loadings of P items on the major genotypic factors in contrast to their small loadings on the principal components of environmental variation is clear if the genotypic factor loadings in Table 1 are compared with their environmental counterparts. The loadings on the first genotypic factor for item 26 (see above) and item 39 ('Would you feel sorry for an animal caught in a trap?') are striking and define this factor virtually exclusively. If this particular genotypic factor were replicated in future studies further evaluation might be warranted. In the light of the discussion of item 26 above it appears that this factor may define a marked difference between the sexes with respect to components of psychoticism. The corresponding loadings on all three environmental factors are small.

The second genotypic factor has 11 loadings with absolute values greater than 0·2. Five of these are E items, two are N items, one is scored for P and the remainder appear in none of the scales. The pattern of loadings for the E items is not consistent with that used in scoring the items for the E scale. This seems to preclude the interpretation of this factor as a genotypic factor of extraversion. Eysenck & Eysenck (1969) report the resolution of the E phenotype into the two oblique components of sociability and impulsiveness. The apparent difficulty in defining a unitary genotypic extraversion factor is consistent with the interpretation that the observed obliqueness of the component phenotypic factors of E results from the unitary environmental modification of more than one independent genotypic trait.

The third factor has 21 loadings with absolute values greater than 0·2. Thirteen of these are for N items suggesting that a unitary genotypic factor related to neuroticism is defined by this axis. Four of the remaining loadings are given by P items suggesting that the genes which predispose an individual towards neurotic behaviour may also be implicated in certain aspects of psychoticism.

Table 3. *Intercorrelations of the Promax factors for the genetical variation of the PEN*

1·00	−0·21	−0·17	0·25	−0·16	0·19	0·07	0·07	0·03	0·05
	1·00	0·17	−0·09	0·01	0·24	−0·09	0·00	−0·04	−0·07
		1·00	−0·20	0·23	−0·11	0·31	−0·11	0·17	0·18
			1·00	−0·18	0·04	0·04	0·12	−0·10	−0·04
				1·00	−0·09	−0·25	−0·08	0·17	0·19
					1·00	0·06	−0·08	−0·13	−0·03
						1·00	0·04	−0·17	−0·03
							1·00	−0·05	0·01
								1·00	0·31
									1·00

The correlations between the Promax factors (Table 3) are mostly small, which implies that the structure defined by these factors consists of a number of sets of genes which do not show a correlated distribution in the sample and operate independently in the determination of phenotypic variation.

Fig. 2 shows the projections of the relevant item vectors on the plane defined by the first two second-order factors of genotypic variation. The clustering of the N items in the upper left quadrant is marked, but the P items are widely dispersed, often with high loadings, in both dimensions. This suggests that a two-dimensional representation may be necessary for the genotypic variation for psychoticism. There appears to be no axis or item cluster which corresponds to a unitary genotypic extraversion factor.

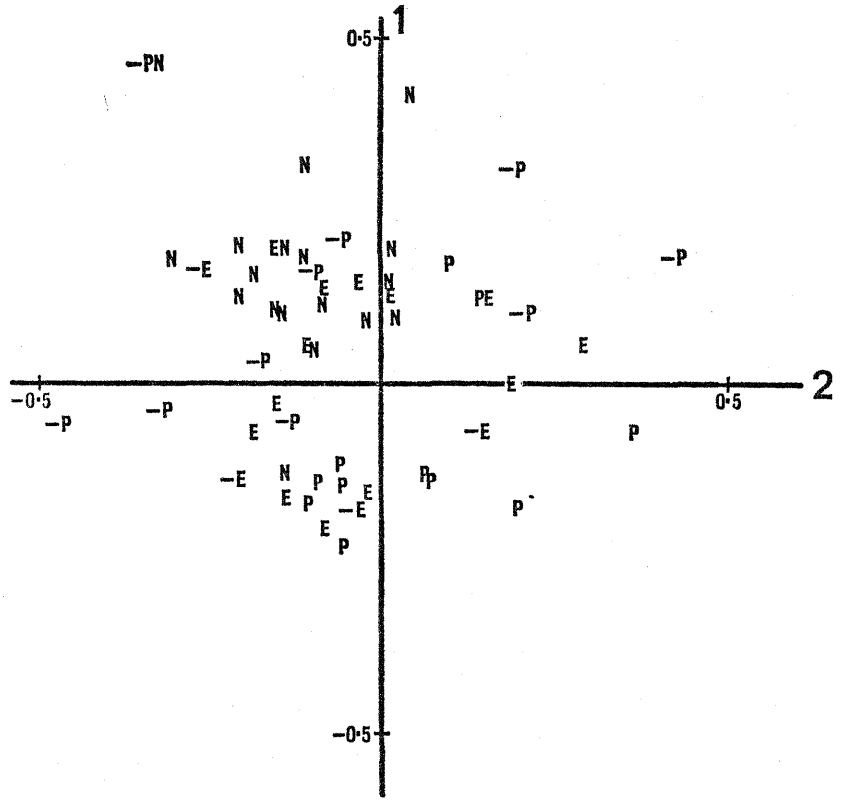

Fig. 2. The second-order structure of genetical variation obtained by canonical analysis. Only items with at least one loading greater than 0·1 are represented.

CONCLUSION

The genotypic and environmental covariance structure obtained for twin responses to the PEN may be analysed separately provided it is assumed that common environmental influences can be discounted. The environmental influences determining variation for E and N act in a unitary manner, but this analysis does not suggest that the P scores reflect any unitary environmental influence. The P items play a greater part in the determination of genotypic structure, although the item vectors require more than one dimension for adequate representation. A well-defined N factor emerges in the genotypic variation, suggesting that neuroticism represents the unitary action of genetical influences. The E items do not form a single genotypic factor, possibly because the observed phenotypic unity of E depends on the correlated environmental modification of more than one underlying genotypic factor. The correlation between P and N reported by the authors of the scale is shown to reflect the correlated influences of both genotype and environment on these factors.

This work, which was carried out whilst the author was in receipt of an S.R.C. research studentship, is part of a project in psychogenetics supported by the M.R.C. The author is indebted particularly to Dr John S. Price for making the data available and to Professors J. L. Jinks and P. L. Broadhurst for their helpful interest. All the computations were conducted on the Birmingham University KDF-9 computer.

REFERENCES

BARTLETT, M. S. (1950). Tests of significance in factor analysis. *Br. J. statist. Psychol.* **3**, 77–85.

BOCK, R. D. & VANDENBERG, S. G. (1968). Components of heritable variation in mental test scores. In S. G. Vandenberg (ed.), *Progress in Human Behaviour Genetics*. Baltimore: Johns Hopkins.

CATTELL, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *J. educ. Psychol.* **54**, 1–22.

CEDERLÖF, R., FRIBERG, L., JONSONN, E. & KAIJ, L. (1961). Studies on similarity diagnosis in twins with the aid of mailed questionnaires. *Acta genet. (Basel)* **11**, 338–362.

EAVES, L. J. (1970). Aspects of human psychogenetics. (Unpublished Ph.D. thesis, University of Birmingham.)

EYSENCK, H. J. & EYSENCK, S. B. G. (1968). A factorial study of psychoticism as a dimension of personality. *Multiv. behav. Res.* **3** (special issue), 15–31.

EYSENCK, H. J. & EYSENCK, S. B. G. (1969). *Personality Structure and Measurement*. London: Routledge & Kegan Paul.

EYSENCK, H. J. & PRELL, D. B. (1951). The inheritance of neuroticism. *J. ment. Sci.* **97**, 441–465.

EYSENCK, S. B. G. & EYSENCK, H. J. (1968). The measurement of psychoticism: a study of factor stability and reliability. *Br. J. soc. clin. Psychol.* **7**, 286–294.

EYSENCK, S. B. G. & EYSENCK, H. J. (1969). Scores on three personality variables as a function of age, sex and social class. *Br. J. soc. clin. Psychol.* **8**, 69–76.

HOPE, K. (1968). *Methods of Multivariate Analysis*. London: University of London Press.

JINKS, J. L. & FULKER, D. W. (1970). A comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behaviour. *Psychol. Bull.* **73**, 311–349.

LOEHLIN, J. C. (1965). A heredity–environment analysis of personality inventory data. In S. G. Vandenberg (ed.), *Methods and Goals in Human Behaviour Genetics*. New York: Academic Press.

LOEHLIN, J. C. & VANDENBERG, S. G. (1968). Genetic and environmental components in the covariation of cognitive abilities: an additive model. In S. G. Vandenberg (ed.), *Progress in Human Behaviour Genetics*. Baltimore: Johns Hopkins.

MATHER, K. & JINKS, J. L. (1971). *Biometrical Genetics: The Study of Continuous Variation*. London: Chapman & Hall.

MAXWELL, A. E. (1961). Canonical variate analysis when the variables are dichotomous. *Educ. psychol. Measur.* **21**, 259–271.

ROUDABUSH, G. R. (1968). Analysing dyadic relationships. In S. G. Vandenberg (ed.), *Progress in Human Behaviour Genetics*. Baltimore: Johns Hopkins.

SHIELDS, J. (1962). *Monozygotic Twins Brought Up Apart and Brought Up Together*. London: Oxford University Press.

Manuscript received 2 December 1971

Revised manuscript received 15 March 1972