Journal of Personality and Social Psychology
1989, Vol. 56, No. 3, 446–455

0022-3514/89/$00.75

Rater Bias in the EASI Temperament Scales: A Twin Study

M. C. Neale
Department of Human Genetics
Medical College of Virginia

J. Stevenson
University of Surrey
Guildford, Surrey, England

Under trait theory, ratings may be modeled as a function of the temperament of the child and the bias of the rater. Two linear structural equation models are described, one for mutual self- and partner ratings, and one for multiple ratings of related individuals. Application of the first model to EASI temperament data collected from spouses rating each other shows moderate agreement between raters and little rating bias. Spouse pairs agree moderately when rating their twin children, but there is significant rater bias, with greater bias for monozygotic than for dizygotic twins. MLE's of heritability are approximately .5 for all temperament scales with no common environmental variance. Results are discussed with reference to trait validity, the person–situation debate, halo effects, and stereotyping. Questionnaire development using ratings on family members permits increased rater agreement and reduced rater bias.

In trying to establish the origins of individual differences in temperament and personality, the family provides an important and unique source of information. Within this setting, family members are able to observe one another's behavior across extended time periods and a wide variety of situations. However, there are a number of issues concerning the accuracy of measurement in such a setting that must be considered before substantive conclusions can be drawn about the influences on individual differences.

A major difference between personality measurement in adults and in young children is that adult personality measurement is usually based on a self-report questionnaire, and juvenile personality is typically assessed by another rater, often a parent. Each of these methods of personality assessment has measurement difficulties; some of these problems are common to both forms of measurement, and others are specific to one or the other. In family studies of personality, the associations between these two types of measure are obtained and compared. The use of twins or adoptees allows the estimation of genetic and environmental influences on individual differences. Before such data can be interpreted, the limitations posed by the measurement techniques need to be established. The aim of this paper is to demonstrate how data from families containing twin children can be used to quantify some of the influences on personality measures and consequently can provide more sensitive and complete estimates of the influences on individual differences in personality.

The accuracy of measures of personality has recently been reviewed by Funder (1987). He has argued for the need to establish a systematic account of social judgments in everyday situations outside the laboratory. In this context, the emphasis should be on whether judges agree with one another rather than on sources of error in social judgments, as is more often the case in social psychological investigation. Funder went on to suggest that as long as subjects are well known to each other, one can obtain at least modest degrees of agreement between self-report and ratings by another. The prevailing theoretical accounts of personality have tended to make strong arguments for the salience of situational factors (e.g., Bem & Allen, 1974; Mischel, 1968), traits (e.g., Cattell, 1982; Eysenck, 1967; McCrae, 1982), or situation–trait interactions (e.g., Epstein, 1983). It is clear that any satisfactory theory will have to take into account each of these sources of influence on an individual's behavior in any one setting (Pervin, 1985). The case for predictive validity of traits is particularly strong when measurements are aggregated (Rushton, Brainerd, & Pressley, 1983), or when some measure of consistency of the trait within the individual is incorporated (Kenrick & Stringfield, 1980). The contrast between the situationist and trait positions has been highlighted recently in the debate about the nature of temperament differences in children (Goldsmith et al., 1987). Rowe (1987) has argued that research designs that allow a separation of genetic and environmental influences can help to resolve some of the issues in the person–situation debate. However, before family or twin data, or both, can be used to resolve the issues surrounding situational and person-centered influences on behavior, important measurement issues need to be addressed.

There are numerous general problems with the use of self-report and rating scales, including response biases, ambiguous items, faking, and acquiescence. Generally, these problems lead to reduced correlation with external validating measures, provided that the external measures are not subject to the same sources of systematic bias. These difficulties have been stressed by Nisbett and Wilson (1977); nevertheless, a case for self-report data has been made by Averill (1983). With subtle questionnaire design and low motivation of volunteer samples to present a favorable image, these difficulties can be minimized (Cronbach, 1970). Self-ratings of personality are potentially subject to a variety of sources of inaccuracy associated with introspection. When responding to questionnaire items, the subject relies on his or her self-concept, which may be inaccu-

Correspondence concerning this article should be addressed to M. C. Neale, Department of Human Genetics, Medical College of Virginia, Box 33, Richmond, Virginia 23298.

rate for a number of reasons. One popular school of thought, known as symbolic interactionism (Schrauger & Schoeneman, 1979) has suggested that the idea of the self is built up as a reflection of the way one appears to others. If the behavior of others toward the self were not consistent, an essentially random self-concept would emerge, which would fail to correlate with other ratings or behavioral or physiological variables. Even if the behavior of others toward the self were consistent, the looking-glass self-image could be inaccurate for several reasons. First, the sample of individuals from whom feedback information is obtained might not be a representative sample of the population. Second, the perception of the attitudes and responses of others may itself be inaccurate. Third, the storage and recall of the self-image may be subject to error. Any of these factors would reduce the correlation between the self-report and external validating measures. One systematic attempt to analyze components of bias in self-report measures is given by Paulhus (1986). He distinguished between self-deception, which he sees as an inevitable component of some personality dimensions, and impression management, which is regarded as a genuine source of measurement error. Data from family studies with twin children enable the assessment of the relative importance of these sources of measurement error on the accuracy of rating scales.

The precision of rating scales is critically dependent on a number of methodological assumptions. First, the rater must know the subject well, as ignorance will inflate error of measurement and reduce correlation with external criteria. Second, the judge's response style (e.g., tendency to view subjects favorably or unfavorably when in doubt) may lead to spuriously high correlations between ratings of different subjects while attenuating agreement between judges. Similar effects are expected if the judge is comparing the behavior of the person being rated either against his or her own self-concept or against some possibly inaccurate general impression of population norms. Halo effects, in which the rater generalizes across categories of behavior, may increase between-rater agreement if raters detect some of the underlying traits and generalize to items about which they are relatively ignorant. It is also possible that when rating several different subjects, there may be some halo effects of similarity—identical twins, for example, are very similar on a number of characteristics (e.g., physical), and the partially ignorant observer may assume similarity where none actually exists. This has considerable importance for the general approach of using ratings of twins to partition genetic and environmental variation. Any tendency for rater bias to be greater in monozygotic (MZ) twins than in dizygotic (DZ) twins would inflate estimates of heritability.

The issue of rater bias in temperament measurement has been addressed recently by Lyon and Plomin (1981) and by Stevenson and Fielding (1985). Both research groups collected data from the parents of twins using both adult and child forms of the EASI (Emotionality, Activity, Sociability, and Impulsivity) temperament scales (A. H. Buss and Plomin, 1975). Parents were required to rate themselves, their spouse, and their twin children. In both these studies, analyses were based on correlations between parental and child temperament measures. Both studies showed there to be no evidence of projection in the ratings by parents of their children's temperament; that is, they did not bias their reports in the direction of making their children appear like themselves. However, there was no attempt to establish the fit of specific models to account for inaccuracy in these ratings. To extend Stevenson and Fielding's (1985) findings, we undertook the present analysis using path models with latent variables to test explicit models of sources of parent agreement and disagreement in their ratings of temperament.

Method

Path Models

The method of path analysis (Wright, 1934) permits the specification of theory in a linear model, relating both observed and unobserved (or latent) variables. Application of this method gives expected correlations between variables, which may be compared with observed correlations collected from a suitable population. This approach is currently popular in the specification of genetic and environmental models of individual differences. Here it is used to specify a model of trait theory in a formal fashion and to allow for the estimation of effects due to rater bias. Multiple rater observations have been used by Heath et al. (1985) to model bias in twins rating themselves and their parents on educational attainment. In the present study, ratings were made by the parents on themselves, their spouses, and their twin children, so some modification of the model presented by Heath et al. is necessary.

In principle, data collected from twins and their parents provide the information to estimate parameters reflecting additive genetic variation, environmental effects shared by twins, cultural transmission from parent to child, and assortative mating (Eaves, Last, Young, & Martin, 1978; Fulker, 1981). In the present case, the resemblance between parent and child may be affected not only by the magnitude of genetic and cultural transmission, but also by a number of effects assumed to be absent. These effects include genetic and environmental nonadditivity, Genotype × Age interactions, and failure of the equal environment assumption of the classical twin study. Perhaps the most critical assumption is that the same phenotype is measured by the juvenile and adult forms of the questionnaire. To avoid the use of these assumptions, we examine the data as two subsets: one in which the parents rate each other and one in which the parents rate their twin children. Using two separate models of familial resemblance and rater bias also helps to emphasize the difference between the type of biases estimated in the two designs. In both cases, path models are used in standardized form to express the correlational structure of the data. Constraints are imposed on the path coefficients in order to keep estimates in the range from -1 to 1. Because standard deviations may be computed for each phenotype, these are also estimated in the model-fitting procedures. In the case of mutual ratings by spouses, these parameters are designated SD_{ij}, so that, for example, SD_{mf} is the standard deviation of mothers' ratings of fathers.

First, consider the case of husband and wife who rate themselves and each other. We assume that agreement between the husband's self-rating and the rating made of him by his wife occurs because both are caused by the husband's underlying temperament (or "latent phenotype"). We make the same assumption for the two ratings of the wife's personality. The two latent phenotypes of the marital pairs are allowed to correlate, reflecting any effects of assortative mating or regional stratification. In addition, we allow for direct effects of the rater's latent phenotype on the rating that he or she makes of his or her spouse. This causal path would be expected to be nonzero if certain sorts of bias were present when people make judgments of others' personality. For example, if having a high score on a test of emotionality led one to perceive others as emotional, then one would expect to find a positive estimate of the path from latent phenotype of emotionality to the rating made of others. Conversely, if having a high emotionality score led one to perceive others

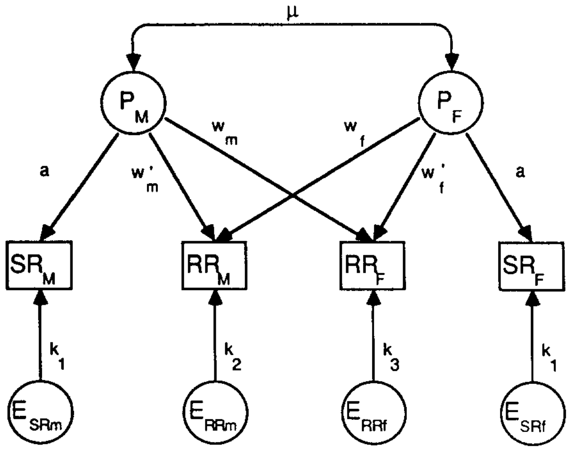

Figure 1. Path diagram showing correlated latent phenotypes of a mother and a father (P_M and P_F), their ratings of their spouses (RR_M and RR_F), and their observed self-ratings (SR_M and SR_F). (E_{RRm}, E_{RRf}, E_{SRm}, and E_{SRf} are uncorrelated residual error variables.)

as relatively unemotional, then a negative estimate of the bias path would be expected. The remaining variance of both self-ratings and ratings made by the spouse also have their interpretations. Residual variance of self-ratings reflects the inaccuracy of the self-rating process. High values indicate the failure of self-ratings to agree with ratings made either on or by the spouse, so that biases associated with the process of introspection will increase this proportion of variance. It is useful to summarize the model for mutual ratings by relatives in a path diagram (see Figure 1). Path diagrams are simply pictorial representations of structural equation models, so we may write this equation for a self-rating that is a function of the latent phenotype and residual variance:

$$SR_M = aP_M + k_1 E_{SR_M},$$

where SR_M, P_M, and E_{SR_M} are the observed self-rating, the latent phenotype, and the residual error component, respectively. A similar expression may be written for the husband's self-rating. The wife's rating of the husband's personality is a function of three variables: the latent phenotype of the husband, the latent phenotype of the wife, and residual error. The structural equation is written

$$RR_M = w'_m P_M + w_f P_F + k_2 E_{RR_M}$$

Because the values estimated for the paths w_m and w_f may differ, as may those for w'_m and w'_f, the residual error path for ratings made by the spouse may differ between the sexes. Although it would be of interest to apply a fully sex-limited version of this model, marital pairs invariably consist of one man and one woman, thus preventing full separation of sex-associated effects. In larger and more varied groups of individuals rating one another, it is possible to specify more elaborate models of rater bias.

The second model defined in this article has two objectives. The first is to obtain maximum likelihood estimates of genetic and environmental components of the most accurate rating of temperament in the juvenile twins. There are numerous ways to estimate genetic and environmental parameters from data collected from twins reared together. In view of the low power to detect dominant genetic effects in the classical twin study (Martin, Eaves, Kearsey, & Davies, 1978), we assume here a model of additive genetic, common, and specific environment effects. In structural equation terms, we may write that the phenotype of the twins is the sum of the effects of the genotype and the environment:

$$PT_1 = hGT_1 + eET_1$$

We partition the environmental variation into two sources: *common environment* (CE), which reflects the effects of environmental factors shared by the twins, and *specific environment* (SE), which reflects the effects of unique individual experiences not shared by the twins. Hence we may write a structural equation for the environment as follows:

$$ET_1 = \beta CE + \gamma SE_1$$

By separating the same-sex twin pairs into groups according to sex and zygosity (MZ male, MZ female, DZ male, and DZ female), we may estimate genetic and environmental parameters specific to each sex. The presence of opposite-sex pairs allows the specification of nonscalar sex-limitation (Eaves, 1977), in which either the genetic or the common environmental components are not perfectly correlated across the sexes. The absence of MZ opposite-sex pairs prevents the simultaneous estimation of nonscalar sex-limited effects in both genetic and environmental variation, which would seem to be a major shortcoming for the use of twins to detect sex-associated variation. However, for many personality traits there appears to be little common environmental variation (Eaves & Eysenck, 1976; Eaves et al., 1978, Fulker, 1981; Goldsmith, 1983; Henderson, 1982; Hewitt, 1984; Jinks & Fulker, 1970; Martin & Jardine, 1986), and under these circumstances any nonscalar sex-limitation must be associated with genetic sources of variation.

The second objective of the model for ratings of twins is to test the psychometric properties of the scales. This includes testing for the presence of rater bias and estimating the amount of error in judgments of personality made by the twins' parents. Again we build a simple linear structural equation model to represent the putative causes of variability

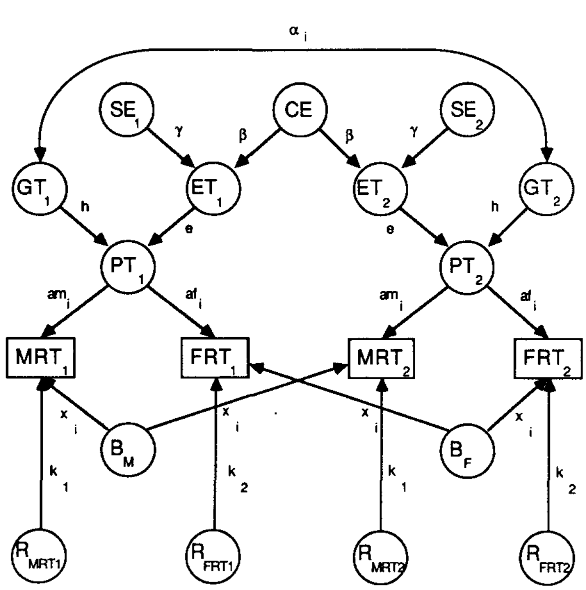

Figure 2. Path diagram showing hypothesised causes of covariation among observed ratings of twins supplied by their mothers (MRT_1 and MRT_2) and their fathers (FRT_1 and FRT_2). (B_M and B_F are latent variables representing projection bias by the parents; PT_1 and PT_2 are latent phenotypes of the twins that have genetic components, GT_1 and GT_2, which correlate [α] 1.0 in MZ twins and 0.5 in DZ twins. In addition, the environments of twins [ET_1 and ET_2] may be correlated due to the common environment [CE]. Residual, specific environmental effects are shown for the latent phenotype of each twin [SE_1 and SE_2]. The four R variables represent residual error variation on each of the measured variables.)

Table 1
Variance–Covariance and Correlation Matrixes of Married Couples for Self- and Spouse Ratings of Emotionality, Activity, Sociability, and Impulsivity Temperament Scales

	Emotionality (df = 528)				Activity (df = 552)			
Rating	1	2	3	4	1	2	3	4
1. RR_F	**45.3797**	−0.2981	23.8851		−4.327221.6494	−1.1000	9.1357	−2.5579
2. SR_M	−0.0065	**47.0230**	−1.8141	25.5432	−0.0635	**13.8508**	−1.6076	7.6496
3. SR_F	0.6047	−0.0451	**34.3797**	0.2423	0.5354	−0.1178	**13.4503**	−1.1269
4. RR_M	−0.0976	0.5657	0.0063	**43.3572**	−0.1441	0.5387	−0.0805	**14.5562**
	Sociability (df = 550)				Impulsivity (df = 548)			
1. RR_F	**19.8822**	1.4932	10.6420	−0.1269	**13.3589**	−0.4613	6.1483	−1.4265
2. SR_M	0.0787	**18.1252**	−0.3296	9.2988	−0.0404	**9.7441**	−0.8430	3.8614
3. SR_F	0.5577	−0.0181	**18.3168**	0.8035	0.5007	−0.0804	**11.2888**	−1.2336
4. RR_M	−0.0076	0.5841	0.0502	**13.9815**	−0.1274	0.4039	−0.1199	**9.3787**

Note. Variances and covariances are given on and above the diagonals (indicated in boldface). Correlations appear below the diagonals. RR_F = mother's rating of father; SR_M = mother's self-rating; SR_F = father's self-rating; RR_M = father's rating of mother.

in the ratings of the twins. The parents are using the same instrument to measure the same individuals at the same age, so a latent variable model is used in which the ratings made by both parents are a linear function of the same underlying trait in the twin (see Figure 2). The strength of the relation between the latent trait and the rating is allowed to differ for the two parents. For example, if fathers are generally more ignorant of the temperament of their children, then the value of a_f will be less than the value of a_m. The parameters x_i reflect the bias of the raters in the ith twin group, so the structural equation corresponding to the rating of the first twin by the mother is

$$MRT_1 = a_{m_i}PT_1 + x_iB_M,$$

and similar equations may be written for the other observed ratings. It is important to recognize that the bias parameter estimated in this second model is a composite of a number of potential sources of bias, and that these differ from those estimated in the model for mutual ratings described previously. When multiple ratings of a single phenotype are made, estimated bias effects subsume (a) the degree of stereotyping, (b) the comparison against the self, (c) the comparison against an idea of population norm, and (d) a type of halo effect from rating people who may be similar on numerous other variables. In the absence of information from a third rater, it is necessary to assume that these effects are of equal degree for the two parents, and that the parents do not correlate in their projection. However, the degree of bias is allowed to differ between twin groups. If the latent trait model is correct, then only the halo effects of multiple rating would give rise to group differences in the values of x_i.

Sample and Measures

Details of the twin sample, zygosity determination, and temperament measures are given in Stevenson and Fielding (1985), so only a brief summary is given here. A total of 939 families with young twins were collected from volunteer sources accessed by the Institute of Psychiatry, UK, and the University of Surrey, UK. Questionnaires were completed by 576 families. Zygosity was determined using a Twin Similarity Questionnaire (Nichols & Bilbro, 1966), and 35 pairs were discarded as no clear zygosity diagnosis could be made. This procedure left a sample consisting of 106 MZ male, 113 MZ female, 129 DZ male, 85 DZ female, and 108 DZ opposite-sex pairs. The twins' mean age was 41.7 months with a standard deviation of 24.8 months. Scores on each of the four temperament scales Emotionality, Activity, Sociability, and Impulsivity were computed using a simple summation procedure as originally described by A. H. Buss and Plomin (1975).

The relative proportions of different types of twin pairs departs somewhat from the usual pattern of an overrepresentation of MZ and female pairs in volunteer twin samples. The current study differed from studies of adult twins because questionnaire response was required by the parents, not by the twins themselves. This method of sampling does not rule out the possibility of bias. However, recent studies (Kendler & Holm, 1985; Lykken, McGue, & Tellegen, 1987; Neale, Eaves, Kendler, & Hewitt, in press) have reevaluated the significance of bias in volunteer twin samples. These studies agree with Martin and Wilson (1982) that differential or bias recruitment into twin samples may have a substantial effect on the estimation of environmental influences on a trait but a less marked one on heritability estimates. One method for detecting whether recruitment biases are likely to affect the results is to test for significant zygosity differences in trait variance. In this study we tested for heterogeneity of variances across all twin groups by fitting a model that constrained variances to be equal across groups but allowed for all correlations to be different. No evidence for heterogeneity was found for the EASI scales (Emotionality, $\chi^2(16) = 6.81$; Activity, $\chi^2(16) = 9.75$; Sociability, $\chi^2(16) = 16.99$; Impulsivity, $\chi^2(16) = 8.89$).

Model Fitting

We calculated variance–covariance matrixes of twin scores corrected for twin age using SPSS[x] (Statistical Package for the Social Sciences, 1983) separately for each variable in the five Sex × Zygosity groups. We calculated variance–covariance matrixes for the spouses by pooling the data from all groups, as we did not expect parents' characteristics to be associated with the zygosity of their twin children. Any family missing data on any single item was discarded from the analysis, thus slightly reducing sample size for each variable. We obtained maximum-likelihood estimates of parameters by minimizing the function

$$F = \sum_{i=1}^{n} df_i\left\{\ln\left[\frac{|\boldsymbol{\Sigma}_i|}{|\mathbf{S}_i|}\right] + \mathrm{tr}(\mathbf{S}_i\,\boldsymbol{\Sigma}_i^{-1}) - p_i\right\},$$

where $\boldsymbol{\Sigma}_i$ and $\mathbf{S}_i$ are the $(p \times p)$ expected and observed matrixes, respectively, corresponding to the ith group; $|\mathbf{S}_i|$ denotes the determinant of matrix $\mathbf{S}_i$; tr denotes the trace of the matrix; and df_i is the degrees of freedom of the ith covariance matrix (Jöreskog, 1969; Neale, Heath, Hewitt, Eaves, & Fulker, in press). In large samples, F is approximately

Table 2
Variance–Covariance and Correlation Matrixes of Pairs of Twins Rated by Their Parents on the Emotionality, Activity, Sociability, and Impulsivity Temperament Scales

	Emotionality				Activity				Sociability				Impulsivity			
Rating	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Male MZ																
		(*df* = 95)				(*df* = 101)				(*df* = 98)				(*df* = 94)		
1. MT_1	**15.82**	5.28	6.72	1.13	**12.47**	8.10	6.65	2.65	**7.14**	4.17	1.72	0.81	**12.88**	5.94	4.30	1.84
2. MT_2	0.37	**12.96**	1.96	4.22	0.57	**15.93**	4.56	7.38	0.65	**5.75**	0.82	2.56	0.49	**11.59**	1.67	4.69
3. FT_1	0.50	0.16	**11.43**	4.11	0.55	0.34	**11.56**	6.50	0.28	0.15	**5.34**	1.70	0.36	0.15	**10.99**	5.37
4. FT_2	0.10	0.40	0.41	**8.70**	0.22	0.55	0.57	**11.27**	0.12	0.43	0.30	**6.11**	0.17	0.46	0.54	**8.92**
Female MZ																
		(*df* = 100)				(*df* = 99)				(*df* = 99)				(*df* = 97)		
1. MT_1	**15.07**	9.60	8.55	5.44	**14.32**	9.86	7.55	4.65	**7.36**	4.10	3.98	1.18	**13.72**	10.53	4.47	5.21
2. MT_2	0.68	**13.32**	5.12	6.82	0.72	**13.04**	3.88	6.23	0.60	**6.27**	2.52	2.50	0.76	**13.85**	1.62	6.32
3. FT_1	0.59	0.37	**14.02**	7.79	0.56	0.30	**12.77**	6.95	0.52	0.36	**7.94**	4.53	0.35	0.13	**12.02**	7.18
4. FT_2	0.43	0.58	0.64	**10.56**	0.37	0.51	0.58	**11.27**	0.18	0.41	0.66	**6.02**	0.42	0.51	0.62	**11.29**
Male DZ																
		(*df* = 117)				(*df* = 120)				(*df* = 117)				(*df* = 113)		
1. MT_1	**13.58**	3.62	6.96	0.36	**17.77**	0.47	7.89	−1.68	**6.81**	−0.15	2.94	−0.82	**13.43**	1.82	5.92	−0.08
2. MT_2	0.26	**14.51**	1.73	6.01	0.03	**15.85**	−2.07	6.70	−0.02	**6.65**	−0.62	2.40	0.12	**16.16**	0.13	8.34
3. FT_1	0.55	0.13	**11.72**	2.94	0.53	−0.15	**12.60**	1.49	0.46	−0.10	**6.08**	1.05	0.50	0.01	**10.41**	3.20
4. FT_2	0.03	0.47	0.26	**11.37**	−0.11	0.48	0.12	**12.33**	−0.13	0.37	0.17	**6.25**	−0.01	0.60	0.29	**11.99**
Female DZ																
		(*df* = 74)				(*df* = 77)				(*df* = 76)				(*df* = 69)		
1. MT_1	**14.17**	0.28	6.45	−1.06	**15.46**	−0.28	9.93	0.25	**5.70**	0.50	1.93	−0.72	**14.64**	2.22	9.03	−1.29
2. MT_2	0.02	**11.30**	−0.05	4.79	−0.02	**13.16**	−3.75	8.25	0.10	**4.79**	0.43	0.96	0.18	**10.50**	−1.24	5.56
3. FT_1	0.53	0.00	**10.46**	1.69	0.67	−0.27	**14.26**	−0.04	0.38	0.09	**4.61**	1.36	0.67	−0.11	**12.34**	0.89
4. FT_2	−0.09	0.48	0.17	**8.93**	0.02	0.62	0.00	**13.39**	−0.15	0.21	0.31	**4.24**	−0.10	0.51	0.08	**11.42**
Male–female DZ																
		(*df* = 94)				(*df* = 95)				(*df* = 97)				(*df* = 61)		
1. MT_1	**14.23**	1.10	5.70	−2.76	**16.29**	−1.11	7.31	−4.09	**6.94**	1.37	2.14	−0.53	**11.59**	2.54	5.43	−1.70
2. MT_2	0.08	**14.15**	−0.03	6.38	−0.07	**15.56**	−3.27	7.99	0.22	**5.76**	−0.39	2.51	0.20	**14.35**	−0.85	6.63
3. FT_1	0.45	0.00	**11.44**	0.01	0.53	−0.24	**11.55**	−1.73	0.34	−0.07	**5.80**	1.00	0.46	−0.06	**12.24**	−1.22
4. FT_2	−0.22	0.50	0.00	**11.40**	−0.30	0.60	−0.15	**11.58**	−0.07	0.38	0.15	**7.45**	−0.15	0.52	−0.10	**11.16**

Note. Variances and covariances are given on and above the diagonals (indicated in boldface). Correlations are given below the diagonals. MT_1 = maternal rating of Twin 1; MT_2 = maternal rating of Twin 2; FT_1 = paternal rating of Twin 1; FT_2 = paternal rating of Twin 2. MZ = monozygotic; DZ = dizygotic.

distributed as chi-squared, with (number of statistics − number of free parameters) degrees of freedom. Model fitting was accomplished with programs E04UAF and E04JAF in the Numerical Algorithms Group (NAG) library (NAG, 1984).

Results

Covariance matrixes and their associated degrees of freedom are shown for spousal ratings in Table 1. Table 2 shows the same statistics for the ratings of the twin children.

Spouse Ratings

The results of fitting the full model for ratings of self and spouse are shown in Table 3. As there are 10 free parameters in the model and 10 observed statistics, the model should fit perfectly to the data, provided that none of the constraints of the model is active. The fit is not perfect for the Emotionality and Sociability scales. For Emotionality this departure is small, whereas Sociability shows significant deviation from the predictions of the model. Inspection of the data matrixes (Table 1) reveals that for Sociability both self–self (r_{SR_F,SR_M}) and spouse–spouse (r_{RR_F,RR_M}) correlations are negative, whereas the two self–spouse correlations (r_{SR_M,RR_F} and r_{SR_F,SR_M}) are positive. The residual covariance matrix (not shown) makes it clear that this pattern of observed correlations is not consistent with the model. As the parameter w_f' is fixed at its boundary value of 1.0, the function value may be interpreted as a test (χ^2 with 1 df) of the hypothesis that the data do not show a significant departure from the model specifications. This test is significant for the Sociability scale; however, the chi-square test is very powerful with such large sample sizes. The discrepancy between the observed and expected statistics is very small.

Table 3
Parameter Estimates Obtained From Fitting the Path Model Shown in Figure 1 to Data on Emotionality, Activity, Sociability, and Impulsivity Temperament Ratings From Married Couples Who Rated Themselves and Their Spouses

	Scale			
Parameter	Emotionality	Activity	Sociability	Impulsivity
μ	−.23	−.22	−.25	−.14
w_m	.18	.06	.20	−.09
w_m'	1.00	.72	.92	.64
w_f	−.07	−.09	.00	−.05
w_f'	1.00	.75	1.00	.51
a	.60	.73	.62	.77
SD_{mf}	6.75	4.65	4.47	3.66
SD_{mm}	6.87	3.72	4.27	3.12
SD_{ff}	5.88	3.66	4.31	3.36
SD_{fm}	6.59	3.82	3.74	3.06
Function value	1.15	0.00	6.29	0.00

Note. Subscripts *m* and *f* stand for *mother* and *father*, respectively. Parameters SD_{ij} represent the standard deviation of the rating by individual *i* on individual *j*. Parameters μ, *w*, *w*′, and *a* are path coefficients represented in the model shown in Figure 1: μ is the correlation between latent phenotypes, *a* is the path from latent phenotype to self-rating, *w* is the path from latent phenotype to rating of spouse, and *w*′ is the path from latent phenotype to the rating of self made by the spouse.

Table 4
Function Values Obtained From Fitting the Mutual Rating Model Shown in Figure 1, Subject to a Variety of Constraints, to the Spouse Data Shown in Table 1

		Scale			
Submodel	*df*	Emotionality	Activity	Sociability	Impulsivity
1 (Full)	—	1.15	0.00	6.29	0.00
2	2	6.76	2.82	7.18	4.12
3	2	5.30	3.45	11.27	2.08
4	1	9.33	8.80	17.17	3.55
5	3	21.69	52.99	28.35	20.26

Note. Submodel 2: $w_m = w_f$; $w_m' = w_f'$. Submodel 3: $w_m = 0$, $w_f = 0$. Submodel 4: $\mu = 0$. Submodel 5: $SD_{mm} = SD_{ff} = SD_{mf} = SD_{fm}$. Subscripts *m* and *f* stand for *mother* and *father*, respectively. Parameters SD_{ij} represent the standard deviation of the rating by individual *i* on individual *j*. Parameters μ, *w*, *w*′, and *a* are path coefficients represented in the model shown in Figure 1: μ is the correlation between latent phenotypes, *a* is the path from latent phenotype to self-rating, *w* is the path from latent phenotype to rating of spouse, and *w*′ is the path from latent phenotype to the rating of self made by the spouse.

Table 4 shows the results of fitting a number of submodels to the marital data. The function values for the submodels may be subtracted from those obtained with the full model, giving likelihood ratio tests (Edwards, 1972) that approximate the chi-squared distribution and therefore allow probability values to be associated with specific hypotheses. Parameter estimates under the full model are the least biased; therefore, we report only goodness-of-fit function values for submodels. It is clear from Submodel 2 that none of the four temperament scales shows any evidence of sex differences in the type or extent of rater bias when the rated person is the spouse. The overall degree of bias is not large as judged from the estimates of w_m and w_f, but for the Emotionality and Sociability scales this bias is significant (Submodel 3). The bias parameter is positive, and is higher for mothers than for fathers, implying a comparative process such that spouses are seen to be more similar to oneself than is actually the case. A consistent feature of the model-fitting results is a negative estimate of μ, the correlation between the spouses' latent phenotypes. Although these estimates are not large, it is unusual that they are negative, because assortative mating appears to be low but positive for a number of personality variables (D. M. Buss, 1984). The function values obtained when the assortative mating parameter is fixed to be zero are shown under Submodel 4 in Table 4. The likelihood ratio tests indicate that the values of μ are significant for all four temperament scales. Note that the parameter estimates of μ are larger than the observed correlations between the self-ratings of the spouses because the expected correlation corresponding to these data points is μa^2, reflecting the inaccuracy of the self-rating procedure. This is an important result, because if the latent phenotypes of spouses are more highly correlated than would appear from self-ratings alone, incorrect conclusions about the genetic resemblance due to assortative mating may be drawn.

In Submodel 5, the standard deviation parameters were constrained to be equal, resulting in a highly significant increase in the function value for all scales of the EASI. Examining the data, it is clear that there is a consistent tendency for the ratings made by the mothers to have a larger variance, particularly when rating their husbands. This may be due in part to the higher degree of bias observed for ratings by women, or due to the husband having less knowledge of the wife's phenotype than vice versa. The latter hypothesis is in accordance with Weiss's (1979) results, in which systematic reduction in information led to reduced variation in personality ratings. Unfortunately, it is not possible to resolve these effects without data from same-sex couples, which may differ from heterosexual couples for other reasons.

Ratings of Twins by Parents

Variance–covariance matrixes for the parental ratings of the twins are shown in Table 2 for each of the sex–zygosity groups. Parameter estimates obtained from fitting the model of external rater bias to these data on twins are shown in Table 5. Under the assumption of multivariate normality, the function values obtained approximate the chi-square distribution, with 26 degrees of freedom. Further function values for submodels that test specific hypotheses are shown in Table 6. The model fits the data on Emotionality very well. Submodel 2 in Table 6 tests for covariation between twins; the difference chi-square between this and the full model is highly significant and therefore indicates that twins correlate for the Emotionality scale. Submodel 3 indicates that the effect of the shared environment does not contribute to the covariation in twins to any significant degree. However, removing heritable effects from the model (Submodel 4) does lead to a significant deterioration in fit. Submodel 5 shows evidence for sex differences in sources of variation: parameter estimates indicate heritability for girls to be higher than for boys and no apparent correlation between genetic effects in boys and girls.

Submodels 6–9 show the results of testing specific hypotheses about the parameters associated with the effects of bias. First is a test of equality of variance, which is nonsignificant, indicating that reporting style does not differ between parents in this respect for the Emotionality scale. Submodel 7 shows that the impact of the latent phenotype on the ratings does not differ according to which parent is the rater, indicating that parents do not differ in the accuracy of their ratings. The x_i parameter estimates from the full model suggest that there is somewhat more projected similarity for monozygotic twin pairs than for dizygotic pairs. Fixing the amount of projection bias to be equal regardless of twin zygosity group (Submodel 8) leads to a significant deterioration in fit. A test of the overall significance of the amount of projection bias is given by Submodel 9, in which all the projection bias parameters are fixed at zero; this gives a very highly significant loss of fit. This is to be expected from inspection of the observed covariance matrixes (Table 2) as the MZ twin correlation exceeds the correlation between the two ratings of an individual in most cases. In addition, the twin correlations across raters are particularly low and are negative in the dizygotic female and opposite-sex twin groups.

Generally, we obtained similar results of model fitting for all four temperament scales. Therefore, discussion of the Activity, Sociability, and Impulsivity scales is brief. In contrast to the Emotionality scale, the full model did not give a good fit to the data on the other scales. It is necessary to consider the ways in which data may depart from the expectations of the model. First, the model predicts equal variance for ratings obtained from different twin groups. The heterogeneity of variance tests reported in the Samples and Measures section are nonsignificant, so this possible departure from expectations does not seem to be important for these measures. The model predicts that the correlation between the two ratings of an individual should be the same regardless of whether the individual has been designated as Twin 1 or Twin 2. Inspection of the data matrixes suggests that this prediction is valid: the replicates are very similar. The same conclusion may be drawn about the cross-correlations, which are predicted to be equal regardless of whether the mother is rating Twin 1 and father is rating Twin 2 or vice versa. The reason for the failure of the model to account for the data on Activity, Sociability, and Impulsivity would seem to be the low and frequently negative correlations between dizygotic twins. Alternative models, including effects of genetic dominance, epistasis, or Genotype × Environment

Table 5
Parameter Estimates Obtained From Fitting the Multiple Rater Model Shown in Figure 2 to the Data on Twins Summarized in Table 2

	Scale			
Parameter	Emotionality	Activity	Sociability	Impulsivity
h_m	.56	.62	.67	.61
h_f	.79	.74	.72	.72
g_{mf}	.00	.00	.00	.43
c_m	.00	.00	.00	.00
c_f	.00	.00	.03	.00
SD_{ms}	3.77	4.01	2.57	3.68
SD_{md}	3.70	3.79	2.45	3.67
SD_{fs}	3.34	3.50	2.48	3.33
SD_{fd}	3.35	3.56	2.49	3.44
a_{m_1}	.60	.75	.83	.59
a_{m_2}	.78	.82	.63	.74
a_{m_3}	.75	.59	.46	.61
a_{m_4}	.59	.83	.42	.90
a_{m_5}	.85	.74	.88	.95
a_{f_1}	.78	.72	.44	.71
a_{f_2}	.71	.64	.70	.51
a_{f_3}	.68	.89	.95	.91
a_{f_4}	.93	.78	.91	.70
a_{f_5}	.57	.80	.42	.53
x_1	.50	.56	.55	.60
x_2	.52	.58	.60	.68
x_3	.42	.38	.31	.42
x_4	.30	.29	.42	.43
x_5	.35	.35	.47	.32
Function value	25.49	58.51	47.84	48.71

Note. Subscripts 1–5 refer to monozygotic (MZ) male, MZ female, dizygotic (DZ) male, DZ female, and DZ opposite-sex twin groups, respectively. Parameters are defined as follows: h_m = square root of heritability in males, h_f = square root of heritability in females, g_{mf} = genetic correlation between males and females, c_m = square root of common variation in males, c_f = square root of common environmental variation in females, SD_i = standard deviation of rating by mother (m) or father (f) on son (s) or daughter (d), a_{i_j} = path from child's latent phenotype to rating made by Parent i in Twin Group j, and x_i = parent bias in Twin Group i.

Table 6
Function Values Obtained From Fitting the Multiple Rater Model Shown in Figure 2, Subject to a Variety of Constraints, to the Twin Data Shown in Table 2

		Scale			
Submodel	*df*	Emotionality	Activity	Sociability	Impulsivity
1 (Full)	26	25.49	58.51	47.84	48.71
2	31	54.47	79.19	72.62	64.54
3	28	25.49	58.51	47.84	48.71
4	29	50.89	78.11	61.27	58.56
5	29	35.46	71.41	51.48	50.39
6	28	25.68	73.41	48.85	56.97
7	31	28.82	65.20	63.58	58.87
8	30	35.55	83.51	64.18	75.52
9	36	145.16	185.79	135.19	179.23

Note. Submodel 2: $h_m = 0$; $h_f = 0$; $g_{mf} = 1$; $c_m = 0$; $c_f = 0$. Submodel 3: $c_m = 0$; $c_f = 0$. Submodel 4: $h_m = 0$; $h_f = 0$; $g_{mf} = 1$. Submodel 5: $h_m = h_f$; $g_{mf} = 1$; $c_m = c_f$. Submodel 6: $SD_{ms} = SD_{fs}$; $SD_{md} = SD_{fd}$. Submodel 7: $a_{mi} = a_{fi}$. Submodel 8: $x_1 = x_2 = x_3 = x_4 = x_5$. Submodel 9: $a_{mi} = a_{fi}$; $x_i = 0$. Parameters are defined as follows: h_m = square root of heritability in males, h_f = square root of heritability in females, g_{mf} = genetic correlation between males and females, c_m = square root of common environmental variation in males, c_f = square root of common environmental variation in females, SD_i = standard deviation of rating by mother (m) or father (f) on son (s) or daughter (d), a_{i_j} = path from child's latent phenotype to rating made by Parent i in Twin Group j, and x_i = parental bias in Twin Group i.

interaction, would give a superior account of the data, but only models of sibling interaction (e.g., Carey, 1986; Neale, 1985) or parental contrast effects could account for negative correlations between twins.

The results for Activity, Sociability, and Impulsivity are only slightly different than those observed for Emotionality. First, ratings made by the mother have larger variance than those made by the father for the Activity and Impulsivity scales. In theory, this could be due to the rating style of mothers, greater impact of the latent phenotype on the mother ratings, or greater bias. Although it would be of interest to test for differences between parents in the degree of rater bias, it is not possible to do so without data from a third rater. Second, there is no evidence for sex differences in variation in either Sociability or Impulsivity. Third, the accuracy of paternal and maternal ratings is not equal for the Sociability scale.

Discussion

A latent phenotype model of multiple ratings has been applied to data collected from the parents of twins. Low negative assortative mating is observed for the parents' ratings of each other, and there is a small effect of rater bias that is significant for the Emotionality and Sociability scales. To assess agreement, married couples provide a useful source of subjects who have a good knowledge of each other. However, for personality variables, use of parents to assess bias effects has low power because the correlation between spouses is low.

Funder (1987) suggested that when familiarity is high, the degree of agreement between self- and other ratings produces correlations between .3 and .6. The findings here for self- and spouse ratings, where familiarity is presumably very high, give correlations ranging from .40 (for mother's self-rating with father's rating of wife on Impulsivity) to .60 (for father's self-rating with mother's rating of husband on Emotionality). Ratings of offspring by their parents show a similar amount of between-rater agreement. Only 2 of 40 such correlations are below .3, with a minimum and maximum of .21 and .67, respectively. These findings are in line with estimates of agreement between parents for behavioral and emotional measures in a recent meta-analysis by Achenbach, McConaughy, and Howell (1987).

The latent phenotype model gives a good account of the data on Emotionality but fails for the data on Activity, Sociability, and Impulsivity. This failure appears to be associated with low and negative DZ twin correlations that are not predicted by the additive genetic, common, and specific environmental model used here. Estimates of the components of variation for Emotionality show the proportion of the variance associated with additive genetic effects to be 31% for boys and 62% for girls. In addition, genetic variability in the two sexes appears to caused by entirely different factors. The same pattern of lower heritability in boys and low genetic correlation across the sexes is seen for all four scales in the EASI, but the sex differences fail to reach statistical significance for the Sociability scale. Common environmental variance is uniformly nonsignificant. This result is in agreement with results obtained for adult personality measures and reflects the low DZ twin correlations for these variables. The similarity between these results and those found for measures of adult personality is striking. First, there is no evidence of common environmental effects on variation in adult personality (Eaves & Eysenck, 1976; Eaves et al., 1978; Fulker, 1981; Goldsmith, 1983; Henderson, 1982; Hewitt, 1984; Jinks & Fulker, 1970; Martin & Jardine, 1986). Second, for the Neuroticism scale of the adult Eysenck Personality Questionnaire (EPQ; Eysenck & Eysenck, 1975), Martin and Jardine found higher estimates of additive genetic variance in women than in men and a genetic correlation of 0.58 between sexes, significantly different from unity. Furthermore, these authors found no such pattern of sex-associated variation for the Extraversion scale of the EPQ. If adult neuroticism is indexed by childhood measures of Emotionality, Activity, and Impulsivity, and adult extraversion is indexed by childhood Sociability, then the results presented in this article are very close to expectations.

The large rater bias effects seen for all four juvenile temperament variables are cause for concern. Their presence is indicated in the data by the twin correlations across raters that are lower than expected, given the level of rater agreement and the twin correlations within raters. In addition, the MZ twin correlations calculated from a single rater frequently exceed the between-rater agreement for an individual. These biases may reflect genuine problems with the EASI temperament scales, such as large amounts of stereotyping, comparison with the self or opinion-of-population norms, or across subject halo effects. For all four scales, biases are significantly higher for MZ than for DZ twins. This result could be due to parents' preconceived notions of the degree of similarity of MZ and DZ twins, those with DZ twins reporting exaggerated differences between twins or those with MZ twins reporting more similarity than actually exists. If so, this might be detected in cases in which parents are mistaken about the zygosity of their twins (Matheny, 1979;

Scarr, 1968). Generalization from other variables that show marked similarity or contrast (halo effects across persons being rated), which could be detected in a multivariate analysis, would also account for group differences in degree of rater bias. A further possibility is that the single latent trait model is incorrect for these data and that the substantial bias effects are due to twins consistently presenting different and heritable aspects of their phenotype to their mother or father. This latter interpretation would refute the idea that EASI temperament ratings are pure traits. Given these large bias effects, the EASI temperament scales would not seem to be an ideal instrument for the measurement of temperament in young children.

We do not believe the measurement of temperament in children by ratings obtained from parents to be impossible. On the contrary, the work here forms a bench mark with which the characteristics of different scales and even the items within scales (Neale, Rushton & Fulker, 1986) may be compared. Jones (1971) suggested that heritability might be used as a criterion for the construction of psychological tests. The use of multiple raters and related individuals, especially in a genetically informative design, allows a new range of criteria to be used in test construction. These criteria include high between-rater agreement and low rater bias, in addition to factorial purity at the level of the phenotype, genotype, or environment. With such careful construction, variation and covariation in juvenile and adult personality may be explored in detail.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist, 38,* 1145–1160.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review, 81,* 88–104.

Buss, A. H., & Plomin, R. (1975). *A temperament theory of personality development.* New York: Wiley.

Buss, D. M. (1984). Marital assortment for personality dispositions: Assessment with three different data sources. *Behavior Genetics, 14,* 111–123.

Carey, G. (1986). Sibling imitation and contrast effects. *Behavior Genetics, 16,* 319–343.

Cattell, R. B. (1982). *The inheritance of personality and ability: Research methods and findings.* New York: Academic Press.

Cronbach, L. J. (1970). *Essentials of psychological testing.* New York: Harper & Row.

Eaves, L. J. (1977). Inferring the causes of human variation. *Journal of the Royal Statistical Society, 140,* 324–355.

Eaves, L. J., & Eysenck, H. J. (1976). Genotype × Age interaction for neuroticism. *Behavior Genetics, 6,* 359–362.

Eaves, L. J., Last, K. A., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human behavior. *Heredity, 41,* 249–230.

Edwards, A. W. F. (1972). *Likelihood.* Cambridge, England: Cambridge University Press.

Epstein, S. (1983). A research paradigm for the study of personality and the emotions. In M. M. Page (Ed.). *Personality: Current theory and research* (pp. 91–154). Lincoln: University of Nebraska Press.

Eysenck, H. J. (1967). *The biological basis of personality.* Springfield, IL: Charles C Thomas.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: University of London Press.

Fulker, D. W. (1981). The genetic and environmental architecture of psychoticism, extraversion and neuroticism. In H. J. Eysenck (Ed.), *The structure and measurement of intelligence* (pp. 102–132). New York: Springer.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Goldsmith, H. H. (1983). Genetic influences on personality from infancy to adulthood. *Child Development, 54,* 331–355.

Goldsmith, H. H., Buss, A. H., Plomin, R., Roehbart, M. K., Thomas, A., Chess, S., Hinde, R. A., & McCall, R. B. (1987). Roundtable: What is temperament? Four approaches. *Child Development, 58,* 505–529.

Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Sundet, J., Nance, W. E., Corey, L. A., & Magnus, P. (1985). No decline in assortative mating for educational level. *Behavior Genetics, 15,* 349–370.

Henderson, N. D. (1982). Human behavior genetics. *Annual Review of Psychology, 33,* 403–440.

Hewitt, J. K. (1984). Normal components of personality variation. *Journal of Personality and Social Psychology, 47,* 671–675.

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

Jones, M. B. (1971). Heritability as a criterion in the construction of psychological tests. *Psychological Bulletin, 75,* 92–96.

Jöreskog, K. G. (1969). A general approach to maximum likelihood factor analysis. *Psychometrika, 34,* 183–202.

Kendler, K. S., & Holm, N. V. (1985). Differential enrollment in twin registries: Its effect on prevalence and concordance rates and estimates of genetic parameters. *Acta Geneticae Medicae et Gemellologiae, 34,* 125–140.

Kenrick, D. T., & Stringfield, D. O. (1980). Personality traits and the eye of the beholder: Crossing some traditional philosophical boundaries in the search for consistency in all of the people. *Psychological Review, 87,* 88–104.

Lykken, D. T., McGue, M., & Tellegen, A. (1987). Recruitment bias in twin research: The rule of two-thirds reconsidered. *Behavior Genetics, 17,* 343–362.

Lyon, M. E., & Plomin, R. (1981). The measurement of temperament using parental ratings. *Journal of Child Psychology and Psychiatry, 22,* 47–53.

Martin, N. G., Eaves, L. J., Kearsey, M. J., & Davies, P. (1978). The power of the classical twin study. *Heredity, 40,* 97–116.

Martin, N. G., & Jardine, R. (1986). Eysenck's contributions to behaviour genetics. In S. Modgil & C. Modgil (Eds.), *Hans Eysenck: Consensus and controversy* (pp. 13–47). Philadelphia: Falmer Press.

Martin, N. G., & Wilson, R. S. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behavior Genetics, 12,* 1–9.

Matheny, A. P. (1979). Appraisal of parental bias in twin studies. Ascribed zygosity and I.Q. differences in twins. *Acta Geneticae Medicae et Gemellologiae, 28,* 155–160.

McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43,* 293–303.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Neale, M. C. (1985). *Biometrical genetic analysis of human individual differences.* Unpublished doctoral dissertation, University of London, UK.

Neale, M. C., Eaves, L. J., Kendler, K. S., & Hewitt, J. K. (in press). Bias in correlations from truncated samples of relatives. *Behavior Genetics.*

Neale, M. C., Heath, A. C., Hewitt, J. K., Eaves, L. J., & Fulker, D. W. (in press). Fitting genetic models with LISREL: Hypothesis testing. *Behavior Genetics.*

Neale, M. C., Rushton, J. P., & Fulker, D. W. (1986). The heritability of items on the Eysenck Personality Questionnaire. *Personality and Individual Differences, 7,* 771–779.

Nichols, R. C., & Bilbro, W. C. (1966). The diagnosis of twin zygosity. *Acta Geneticae et Medicae Gemellologiae, 16,* 265–275.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84,* 231–279.

Numerical Algorithms Group. (1984). Numerical Algorithms Group FORTRAN Library Manual, Mark 11. Oxford, England: Author.

Paulhus, D. L. (1986). Self-deception and impression management in test response. In A. Angleiter & J. S. Wiggins (Eds.), *Personality assessment via questionnaires.* Berlin, West Germany: Springer-Verlag.

Pervin, L. A. (1985). Personality: Current controversies, issues and directions. *Annual Review of Psychology, 36,* 83–114.

Rowe, D. C. (1987). Resolving the person-situation debate: Invitation to an interdisciplinary dialogue. *American Psychologist, 42,* 218–227.

Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin, 94,* 18–38.

Scarr, S. (1968). Environmental bias in twin studies. *Eugenics Quarterly, 15,* 34–40.

Schrauger, J. S., & Schoeneman, T. J. (1979). Symbolic interactionist view of self-concept: Through the looking glass darkly. *Psychological Bulletin, 86,* 549–573.

Statistical Package for the Social Sciences. (1983). *SPSS^X User's Guide.* London: McGraw-Hill.

Stevenson, J., & Fielding, J. (1985). Ratings of temperament in families of young twins. *British Journal of Developmental Psychology, 3,* 143–152.

Weiss, D. S. (1979). The effects of systematic variations in information on judges' descriptions of personality. *Journal of Personality and Social Psychology, 37,* 2121–2136.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5,* 161–215.

Received July 6, 1987
Revision received December 16, 1987
Accepted July 14, 1988 ■