

# ENVIRONMENTAL VERSUS GENETIC VARIANCE COMPONENT MODELS FOR IDENTICAL TWINS: A CRITIQUE OF JINKS AND FULKER'S REANALYSIS OF THE SHIELDS DATA

**Peter H. Schönemann**

Purdue University, Department of Psychological Sciences  
West Lafayette, IN 47907, U.S.A.

## **Abstract**

*It is shown that the genetic model Jinks and Fulker (1970) fitted to the Shields' (1962) twin data is qualitatively inconsistent with systematic trends in these data and, as a result, produces an inordinately large proportion of negative variance estimates. In contrast, a purely environmental model yields qualitative predictions consistent with the Shields data and admissible parameter estimates throughout. Quantitatively, it fits the Shields data twice as well as Jinks and Fulker's genetic model. Hence their farreaching conclusions are not supported by the Shields data. This reevaluation illustrates that purely descriptive models, even if they were used with circumspection, remain intrinsically inconclusive about nature/nurture questions because the possibility can never be ruled out that other models may fit the same data even better.*

**Key words:**

**Mots clés :**

## 1. INTRODUCTION

Ever since Galton (1883), studies of identical twins have been considered "the most available materials for a scientific examination of the role of nature versus nurture in human psychology" (Morgan, 1929, p. 35). The still widely held belief that many mental traits, including "intelligence", are inherited to a considerable degree (until quite recently, the figure was 80%), can be traced in large part to twin studies. These conclusions should not be taken lightly: Between 1921 and 1964, 33,374 US citizens were sterilized on the grounds that they were "mentally retarded", that is, because they scored below 70 on so-called "intelligence tests" (Robitscher, 1973, p. 123).

Compared with the scope of its social and ethical implications, the empirical basis for twin research on mental abilities to this day is actually quite narrow, because "twins are uncommon, and monozygotic twins who have been separated from one another in early childhood are of great rarity" (Slater, in Shields, 1962, p. vii). In an effort to broaden this base, Shields appealed in the late 50's in a BBC television program for volunteers among monozygotic and dizygotic twins to submit to a series of mental tests. He published his results in Shields (1962). Since its publication, this study has occupied a pivotal position in the still ongoing debate about the degree of inheritance of mental traits. The Shields data have been analyzed and reanalyzed by numerous other investigators, including Jinks and Fulker (1970) and, more recently, Farber (1981).

This particular data set acquired still more empirical weight when Kamin (1974) discovered that almost half of the available data base of monozygotic twins raised apart (MZAs), the 53 pairs reported by Burt, had to be discounted as untrustworthy. Prior to this discovery, the Shields data contributed 44/137 pairs (= 32%) to this data base, thereafter more than half (44/84 pairs = 52%).

Shields recorded the values of various physical and four psychological variables: "After discussion with Professor Eysenck it was decided to use the Dominoes Intelligence Test and the Synonyms section (Set A) of the Mill Hill Vocabulary Scale (Form B, 1948)" (Shields, 1962, p. 58) as the two cognitive variables. Two personality variables were derived from a "Self-Rating Questionnaire (SRQ) ... especially designed for us by Professor H. J. Eysenck ... so as to yield the greatest amount of information compatible with its shortness" (*ibid.*, p. 65). This questionnaire consisted of 38 items. "Twenty-two items contribute to a score for extraversion and twenty-two items towards a score for neuroticism, six of the items being common to both dimensions" (*ibid.*, p. 66). Shields' sample comprised 44 pairs of monozygotic twins raised apart (MZAs), 44 pairs of monozygotic twins raised together (MZTs), and 32 pairs of dizygotic twins raised together (DZTs).

In his (1962) book, Shields presented these data in considerable detail. After analyzing them in various ways, he stated his final conclusions with noteworthy caution: "The above discussion of genetical and environmental causes of variation has suggested that on both sides these are multiple and interacting ... The existence of twins who are alike in all their genes has nevertheless enabled us, by means of appropriate comparisons, to demonstrate something of the importance of heredity for a wide variety of personal characteristics. The very fact that such twins differ, sometimes extensively, is of itself evidence of non-genetical effects. To some it may come as a surprise that twins brought up together differ so much." (loc. cit., p. 156).

In contrast, when Jinks and Fulker announced their results of a reanalysis of these data in Jinks and Fulker (1970), they were much more specific in their conclusions than Shields had been:

(a) The authors claimed to have resolved the nagging problem of genetic-environmental interactions, and of correlated environments of MZAs which, if present, would invalidate any inferences about the presumed heritability of psychological traits: "... the inheritance of most of the psychological measures reanalyzed conform to a simple model. In view of pessimism, over the possible influence of correlated environments and genotype-environment interactions ... it is reassuring to find that they are by no means universal phenomena" (p. 347);

(b) They claimed to have established uniformly "high heritabilities recorded in Table 30" (p. 347) for the four measures employed by Shields;

(c) They claimed that, "For the section Neuroticism, assortative mating was indicated and only additive gene action" (p. 348);

(d) and, further, that "IQ showed strong directional dominance for high expression" (p. 348);

(e) In spite of the small sample sizes and the abbreviated tests, the authors claimed to be able to estimate the number of genes controlling IQ: "the high number of genes estimated to be controlling IQ (>20 and approximately 100) fully confirms that this trait is under polygenic control" (p. 348).

This study has attracted considerable attention. The reactions of psychologists ranged from favorable to enthusiastic:

Loehlin, Lindzey, and Spuhler (1975, p. 289): "For more sophisticated methods of heritability coefficients and variance components, the reader is referred to Jinks and Fulker (1970)".

Jensen (1980, p. 184): "Genetical models that are based on general principles of genetics and are applicable to metric traits in all plants and animals fit the various kinship correlations for IQ remarkably well (Jensen, 1973a; Jinks & Fulker, 1970)".

Eysenck (1973, p. 262): "The Jinks and Fulker paper ... is the cornerstone on which any future argument about heritability must be based".

By comparison, geneticists were more reserved:

Vetta and Hirsch questioned claim (a) that genotype does not interact with environment, on technical grounds: "Analysis of this work by Dr. Vetta revealed a mistake in the algebra on p. 314 of the Jinks and Fulker article. ... Even though Professor Jinks accepted Dr. Vetta's correction of his algebraic error, Richard Hernstein, editor of the psychology journal in which it appeared, refused to allow Dr. Vetta to publish his correction." (Hirsch, 1981, p. 23, see also Hirsch, 1990).

King (1981) expressed some doubts about the precision of the numerical estimates Jinks and Fulker had reported, usually to four decimal places for samples ranging from 11 to 29 pairs: "One cannot help wondering whether the claims to precision of the estimates turned out by the kind of ponderous statistical machinery at work in Jinks and Fulker's 1970 paper are not spurious" (p. 86). King also drew attention to a more critical problem which will figure centrally in the present paper: After comparing the heritability estimates typically reported for mental traits (e.g., around .80 for IQ) with those found in controlled animal experiments (e.g., milk production in cattle: .30, staple length of wool in sheep: .25, egg production in poultry:  $\leq .50$ , p.79), King concluded: "In estimating heritability of characters in domestic animals the breeders make every effort to randomize environmental factors [while] studies of heritability of human intelligence are probably more deficient in failing to randomize environmental factors" (p. 82).

More generally, the point of the present paper is to assess the stringency of some of the reasoning which led Jinks and Fulker to their conclusions, and also to assess the tightness of fit of their genetic model to the Shields twin data. The relevant subset of the Shields data is reproduced in Table 1. The sample sizes are given by the within degrees of freedom. The between and within sums of squares have been recomputed from the various Tables in the Jinks and Fulker (1970) article and were checked against the total variances which the authors also reported. Since there are not enough data for the two cognitive variables on the dizygotic samples, the present reanalyses will be limited to identical twins. Since Jinks and Fulker (1970) did not report the neuroticism values for the male MZT sample (p. 326), these values were computed directly from the Shields (1962) report. In this connection, a minor discrepancy with the values given by Jinks and Fulker for extraversion was found and corrected.

## 2. THE GENETIC MODEL BY JINKS AND FULKER

To account for the Shields data, Jinks and Fulker postulate a variance components model which provides for four mutually uncorrelated latent variables to

explain the observed variables  $y$  (neuroticism, extraversion, Mill Hill, and Dominoes test). The structural part of the Jinks and Fulker model is:

TABLE 1. Between and within sums of squares for the Shields data recomputed from Jinks and Fulker (1970).

Variable	Sex	Twin Type	Between/ Within	df	SS	MS
Neuroticism	male	together	B	13	104.61	8.05
			W	14	109.62	7.83
		apart	B	13	382.98	29.46
			W	14	70.00	5.00
	female	together	B	28	620.48	22.16
			W	29	235.48	8.12
		apart	B	25	728.00	29.12
			W	26	251.16	9.66
Extraversion	male	together	B	13	278.96	21.46
			W	14	92.25	6.59
		apart	B	13	293.43	22.57
			W	14	50.00	3.57
	female	together	B	28	405.35	14.48
			W	29	188.13	6.49
		apart	B	25	766.18	30.65
			W	26	200.00	7.69
Mill Hill	male	together	B	11	189.46	17.22
			W	12	36.48	3.04
		apart	B	14	962.87	68.78
			W	15	159.50	10.63
	female	together	B	23	796.67	34.64
			W	24	111.00	4.63
		apart	B	24	1372.00	57.17
			W	25	178.50	7.14
Dominoes	male	together	B	10	905.09	90.51
			W	11	54.01	4.91
		apart	B	13	1423.61	109.51
			W	14	270.50	19.32
	female	together	B	22	2556.40	116.20
			W	23	571.00	24.83
		apart	B	23	4182.67	181.86
			W	24	462.00	19.25

TABLEAU 1. Sommes des carrés (inter et intra) des données de Shields recalculées d'après Jinks et Fulker (1970).

$$y - \mu = g_b + g_w + e_b + e_w,$$

where  $\mu$  is the population mean of  $y$ ,  $g_b$  a between-family genetic variable,  $g_w$  a within-family genetic variable,  $e_b$  is a between-family environmental variable, and  $e_w$  a within-family environmental variable. The stochastic restrictions are:

$$E(g_b, g_w, e_b, e_w) = \emptyset', \quad \text{Var}(g_b, g_w, e_b, e_w) = \text{diagonal}.$$

Thus, all latent variables are uncorrelated and their means are zero. Under this model, the four latent variables make different contributions to the scores of the MZAs, MZTs, and DZTs:

Since MZTs share the same genes, they both have the same values on the two genetic variables  $g_b$  and  $g_w$ . Since they also grow up in the same environment, they have the same value on the between-family environment variable  $e_b$ . Hence MZTs differ only on the within-family environment variable  $e_w$ . If one denotes the observed score of one twin by  $y_1$ , and that of the other twin by  $y_2$ , the above model accounts for the scores of monozygotic twins raised together (MZTs) as

$$\begin{aligned} \text{MZTs:} \quad y_1 - \mu &= g_b + g_w + e_b + e_{w1} \quad (\text{1st twin}) \\ y_2 - \mu &= g_b + g_w + e_b + e_{w2} \quad (\text{2nd twin}), \end{aligned}$$

where the variables on the right again have zero means and are mutually uncorrelated, and  $\text{var}(e_{w1}) = \text{var}(e_{w2}) = \text{var}(e_w)$ .

In contrast, monozygotic twins raised apart (MZAs) differ also on the between-family environmental variable  $e_b$ , so that their observed scores are modelled as

$$\begin{aligned} \text{MZAs:} \quad y_1 - \mu &= g_b + g_w + e_{b1} + e_{w1} \\ y_2 - \mu &= g_b + g_w + e_{b2} + e_{w2}, \end{aligned}$$

where the six variables on the right have zero means and are mutually uncorrelated, and now, in addition,  $\text{var}(e_{b1}) = \text{var}(e_{b2}) = \text{var}(e_b)$ .

For dizygotic twins raised together (DZTs), this model provides for two different values on the within-family genetic variable  $g_w$ . For dizygotic twins raised apart (DZAs) it provides for different values on  $g_w$ ,  $e_b$ , and  $e_w$ , so that they only share a common value on  $g_b$ . Finally, for unrelated children, all four variables are allowed to take on different values.

Intuitively, this simple model makes some sense as a first approximation if (a) one believes that genetic variables may contribute to a particular test performance, (b) one is willing to assume that the variances of the latent variables remain the same across all four types of twins, (c) all four latent variables are uncorrelated, and (d) no other variables (e.g., gene-environment interactions) enter the observed variables  $y$ . Jinks and Fulker do not give any substantive interpretation for the four latent variables  $g_b$ ,  $g_w$ ,  $e_b$ ,  $e_w$ . However, it is clear that any measurement error must be absorbed by  $e_w$ .

### 3. SOME QUALITATIVE PREDICTIONS OF THE JINKS AND FULKER MODEL

The defining assumptions of the Jinks and Fulker model imply a number of qualitative consequences which can be checked against the data reported by Shields.

In Appendix A1 it is shown that the conventional sums of squares between (SSB) and sums of squares within (SSW) can be computed as half the sum of squared row sums and half the sum of squared row differences from the  $N \times 2$  table of deviation scores around the grand mean. In Appendix A3 these results are used to derive the expected values for these sums of squares, and hence also for the mean squares, from the assumptions of Jinks and Fulker's model.

In Appendix A2 it is further shown how these sums of squares relate to the so-called intra-class correlation, a summary index of within-family, between-twin resemblance which can be defined either in terms of sums of squares, or in terms of mean squares. For pairs it varies between -1 and 1, as does the ordinary correlation coefficient. It is perfectly adequate to interpret an intra-class correlation as an ordinary (product-moment) correlation computed from a "double entry table" which contains each score pair twice, once as  $(y_1, y_2)$  and once as  $(y_2, y_1)$ , so as to remove the ambiguity whether a sib's score is to be entered on the right or left (e.g., Harris, 1913). In passing, note that it is not a variance ratio, as is often said (e.g., Scheffe, 1959, p. 223; Winer, 1971, p. 244) because the numerator is not a variance but a covariance, and hence can be negative. In Appendix A3 it is shown that the genetic model proposed by Jinks and Fulker implies, among other things,

$$(a) E(MSBA + MSWA) = EMSBT + MSWT,$$

where  $E(MSBT)$ ,  $E(MSWT)$  denote the expected values of the between and within sums of squares for the MZT's and  $E(MSBA)$ ,  $E(MSWA)$  those for the MZA's,

$$(b) E(MSBA - MSWA) \leq E(MSBT - MSWT),$$

and, as a consequence,

$$(c) r_A - r_T \leq 0,$$

where  $r_A$  denotes the intraclass correlation (based on mean squares) for MZAs, and  $r_T$  the intraclass correlation for MZTs.

These predictions are reasonable if one believes the observed scores contain a genetic component. In this case, MZTs should be more similar than MZAs, because the former share both the genes and a common environment, while the latter only share the genes, but presumably are exposed to different environments.

However, the actual data reported by Shields (see Table 1) conflict with these predictions, as can be seen from Table 2.

TABLE 2. Ordinal predictions of the genetic and environmental models: Observed and predicted differences between MZAs and MZTs of mean squares and intraclass correlations.

Variable	Sex	(MSBA + MSWA) -(MSBT + MSWT)	(MSBA - MSWA) -(MSBT - MSWT)	$r_A - r_T$
Observed				
Neuroticism	male	18.58	24.24	.70
	female	8.50	5.42	.04
Extraversion	male	-1.91	4.13	.20
	female	17.37	14.97	.22
Mill Hill	male	59.15	43.97	.03
	female	25.04	20.02	.01
Dominoes	male	33.41	4.59	-.20
	female	60.08	71.24	.16
Predicted				
Jinks and Fulker Genetic Model		= 0	≤ 0	≤ 0
Environment Model		≥ 0	≥ 0	≥ 0

Note: If  $\text{var}(e_b) = 0$ , the Jinks and Fulker Model predicts 0 for all 4 comparisons. For the observed intraclass correlations, see Table 4.

TABLEAU 2. Prédiction ordinales des modèles génétique et comportemental : différences des carrés moyens et des corrélations intra-classes observées et prédites entre MZA et MZT.

Specifically, the ordinal predictions (a), (b), and (c) of the genetic model proposed by Jinks and Fulker are violated by the Shields data for prediction

- (a) in 7 out of 8 (= 88%) cases,
- (b) in 8 out of 8 (= 100%) cases, and
- (c) in 7 out of 8 (= 88%) cases.

If, for the sake of argument, we assume approximate independence, then the null-hypothesis that the differences in the first or second row of Table 2 are actually zero implies that the probability of obtaining 7 out of 8 positive differences is given by the binomial as  $(1+2^3)2^{-8} = .035$  for each row, so that the joint outcome in both rows has probability less than .001.

Prediction (a),  $E(\text{MSBA} + \text{MSWA}) = E(\text{MSBT} + \text{MSWT})$ , assumes added significance for Jinks and Fulker's contention that both genotype-environment interactions and correlated environments are negligible. Therefore, the authors



propose to test equality (a) more stringently with an F-test based on the total variances in the MZT and MZA groups. Strictly speaking, their model does not predict that the two total variances are the same, but rather that the two sums of mean squares are the same (see E(MSTA), E(MSTT) in Appendix A4). But this is a minor point of little consequence because the degrees of freedom within and between differ only by 1.

Much more serious is the power problem (see also Wahlsten, 1990). According to the two equations Jinks and Fulker give at the bottom of p. 315 in their (1970) paper, in their notation,

$$\text{MZTs: } \sigma_T^2 = \sigma_{wh}^2 + \sigma_{bh}^2 + \sigma_{we}^2 + \sigma_{be}^2 + 2r_{wh,we} \sigma_{wh} \sigma_{we} + 2r_{bh,be} \sigma_{bh} \sigma_{be}$$

$$\text{MZAs: } \sigma_T^2 = \sigma_{wh}^2 + \sigma_{bh}^2 + \sigma_{we}^2 + \sigma_{be}^2 + 2r_{wh,we} \sigma_{wh} \sigma_{we} + 2r_{bh,we} \sigma_{bh} \sigma_{we}$$

the two total variances they wish to compare are composites of six terms, four postulated variances and two covariances, and differ in only one term, the remaining five are the same for both groups. According to the authors: "Now except under exceptional circumstances of internal balancing, these two  $\sigma_T^2$ 's will not be expected to be equal unless the correlations contribute only an insignificant amount of covariance to the respective total variances" (p. 315).

However, if one assumes for the sake of argument that all 4 postulated variances are equal to  $k$ , and the constant correlation  $r_{wh,we} = .5$ , then one obtains the F-ratio

$$F = \sigma_{T1}^2 / \sigma_{T2}^2 = (5k + 2kr_1) / (5k + 2kr_2) = (1 + 2r_1/5) / (1 + 2r_2/5),$$

which varies between 1 and 1.40 when  $r_i \geq 0$  and between 1 and 2.33 otherwise as a function of the values of  $r_1$ ,  $r_2$ , which are the only two parameters left to vary. The critical values for a .05 level test are around 3 for the smaller groups, and around 2 for the larger, pooled groups. Hence it is virtually impossible to reject the homoscedasticity hypothesis, even if one ignores the added complication that such a stringent alpha level used by Jinks and Fulker further increases the risk of type II errors. For a more comprehensive discussion of the power problem in heritability models with special attention to interaction tests, see the recent BBS target article by Wahlsten (1990).

In any case, the consistent departures of the data from the predictions of the genetic model proposed by Jinks and Fulker cast sufficient doubt on the adequacy of their model that one may wish to look for alternative models which are more compatible with the qualitative aspects of the Shields data.

#### 4. AN ENVIRONMENTAL MODEL FOR THE SHIELDS DATA

To begin with, one may follow up on King's suggestion and question the assumption of uncorrelated environments for the MZAs ( $\text{cov}(e_1, e_2) = 0$ ) which has also been troubling to other observers of the current behavior genetics

scene. Jinks and Fulker mention this crucial assumption only in passing -- "With only  $MZ_T$  and  $MZ_A$  ... the genetic component can still be estimated without bias, provided, of course, that separated twins are randomly assigned across environments" (p. 321) -- and do not comment on its plausibility. As noted earlier, this assumption is absolutely critical for any attempts to separate genetic from environmental contributions to test performance. As King also point out, it seems a rather implausible hypothesis on common-sensical grounds alone: Of course twins are not "randomly assigned" to different environments when they are put up for adoption. A priori it seems more reasonable to expect that well-to-do families, if they have reason to give up a child for adoption, will try to ensure that it is raised in roughly comparable socio-economic surroundings, e.g., by relatives. Less well-to-do families, in turn, will have to be satisfied with an adopting environment similar to their own. Thus, one might postulate an additional parameter  $\rho_{e1,e2}$  to account for correlated environments between  $MZ_A$ s and predict that it will be positive.

However, the inclusion of such a correlation parameter would not resolve inconsistency (a), that  $MSB+MSW$  is larger for the  $MZ_A$ s than the  $MZ_T$ s in 7 out of 8 cases, often much larger. Any change in the covariance assumptions simply redistributes the quantity  $E(MSB+MSW)$  between the  $MZ_A$ s and  $MZ_T$ s under the Jinks and Fulker model, but does not change the prediction of equality between both groups.

In order to predict a larger  $E(MSB+MSW)$  for the  $MZ_A$ s, an additional variable must be added to the structural equation for the  $MZ_A$ s which does not show up in the structural equation for the  $MZ_T$ s. Let this variable be  $s$  (for "separation variable") and assume, as a first approximation, that it takes on the same values for both  $MZ_A$ s of the same family, and thus varies only between families. A possible substantive interpretation of  $s$  might be length of separation (see, e.g., Farber, 1981), although there may be others. Farber, in her thorough reassessment of the quality of the extant  $MZ_A$  data, also noticed the "paradox that twins with the least contact may most frequently be the most alike" (p. 254). For further empirical evidence of this thoughtprovoking finding, see Farber (1981, p. 256, Table 10.1).

To account for the consistent finding that  $r_A > r_T$ , the additional variable  $s$  must also affect the numerator,  $E(MSB-MSW)$ . To see that such a change indeed yields the correct qualitative predictions, consider the ratio

$$r(x) = (a + x)/(a + b + x), a, b, x > 0,$$

which increases with  $x$ . Concretely, if  $a$  is .1 and  $b = 1$ , then  $r(0) = .09$  is its value for  $x = 0$ . If  $x = 1$ , one has  $r(1) = .52$ , if  $x = 2$ ,  $r(2) = .68$ , if  $x = 5$ ,  $r(5) = .84$ , and if  $x = 10$ ,  $r(10) = .91$ . Thus, increasing  $x$  (concretely,  $\text{var}(s)$ ) not only increases the magnitude of the denominator but also the magnitude of the ratio, in agreement with the data reported by Shields.

These purely formal considerations suggest that a more adequate variance component model for the Shields data might be

$$y - \mu = g + e + d + s,$$

where  $g$  denotes the genetic component,  $e$  the environmental component,  $d$  the measurement error, and  $s$  the separation variable needed to account for  $MSBA + MSWA \geq MSBT + MSWT$  and  $r_A \geq r_T$ , and  $s$  enters  $y$  only for the MZAs but not for the MZTs. In response to concerns about possibly correlated environments such a model should further include a correlation parameter  $\rho_{e1e2}$  for the MZAs, raising the total number of parameters to be estimated from three for the Jinks and Fulker model ( $\text{var}(g)$ ,  $\text{var}(e_b)$ ,  $\text{var}(e_w)$ ) to five for such a correlated environments model ( $\text{var}(g)$ ,  $\text{var}(e)$ ,  $\text{var}(d)$ ,  $\text{var}(s)$ ,  $\text{cov}(e_1, e_2)$ ). To reduce this number of parameters by one, it is proposed to eliminate  $\text{var}(g)$  and thus to model the Shields data under the strong assumption that the observed variables contain no genetic component at all. Thus one arrives at a 4-parameter "correlated environments model" (CE4):

$$\text{MZTs: } y_1 - \mu = e + d_1$$

$$y_2 - \mu = e + d_2.$$

$$\text{MZAs: } y_1 - \mu = e_1 + d_1 + s \quad \text{with } \text{cov}(e_1, e_2)$$

$$y_2 - \mu = e_2 + d_2 + s = \rho_{e1e2} \text{var}(e) \neq 0.$$

On fitting this model to the sums of squares of the four data sets pooled over sex, i.e., 8 sums of squares for each of the 4 variables, all variance estimates of the latent variables were positive and the following estimates for the environments correlations for the MZAs were obtained:

$$\text{Neuroticism, } \rho_{e1e2} = .91, \quad \text{Extraversion, } \rho_{e1e2} = .94,$$

$$\text{Mill Hill, } \rho_{e1e2} = .72, \quad \text{Dominoes, } \rho_{e1e2} = 1.04.$$

This outcome suggests that the fit may deteriorate only slightly if one sets  $\rho_{e1e2} = 1$ , to arrive at a 3-parameter correlated environments model (CE3) which can be fitted to each sex group separately. Such a model contains the same number of parameters as the genetic model JFG3 proposed by Jinks and Fulker, so that the relative fit of both models can be compared in each case.

## 5. QUANTITATIVE COMPARISON BETWEEN THE GENETIC AND THE CORRELATED ENVIRONMENTS MODEL

The expected between and within sums of squares predicted by the 3-parameter Jinks and Fulker model (JFG3) are derived in Appendix 4. On fitting this model to the sums of squares (Table 1) by the method of least squares, the least squares estimates (LSEs) shown in Table 3 were obtained. These estimates differ slightly from those reported by Jinks and Fulker, which are also given for comparison. The discrepancies in the estimates are due to (a) some

corrections in the observed sums of squares given by Jinks and Fulker noted earlier, (b) erratic use of the available data base by Jinks and Fulker, who fitted their model to various eclectic combinations of the identical twin and dizygotic twin groups. For example, on p. 326 they discard the male MZTs for the analysis of neuroticism, but include it again on p. 327. In contrast, the present reanalysis is always based on the same four groups for all four variables, and (c) a revised least squares criterion: Jinks and Fulker fitted their model to the mean squares while the present reanalysis is based on sums of squares.

As a check on (c), the mean squares analyses were also performed for the four pooled groups, resulting in only minor changes in the estimates, suggesting that (b) is the main source of the numerical discrepancies in the LSEs.

More striking than the minor discrepancies in the LSEs are the consistencies across the 8 within sex analyses and the 4 pooled analyses. In particular, regardless which least squares method is used, one finds that the estimates for the between-family environment variance component  $\text{var}(e_b)$  turn out negative with stunning regularity in all 12 cases. Jinks and Fulker cope with this surprising finding by constructing confidence intervals around these negative estimates. Since they always include zero, the authors conclude that  $\text{var}(e_b)$  is zero in each case. This inference can be questioned on at least five counts:

(a) Empirically: since  $\text{var}(e_b) = 0$  implies  $r_A = r_T$  under their model, it conflicts with the consistent finding that  $r_A > r_T$  for the Shields data.

(b) Substantively: Once the problem of measurement error is taken into account, such a finding implies that the test performance is entirely due to the genetic component because the intraclass correlations, once corrected for attenuation, are unity. For example, the 8 intraclass correlations for the cognitive variables, the Mill Hill and the Dominoes, average in the high 70s, which roughly equals the parallel form reliabilities of "Cattell's Culture Fair Intelligence Test" for 2 white groups of the data reported by Osborne (1980). So far as I know, not even the most extreme hereditarians went so far as to claim that the performance on mental tests is entirely due to genetic causes.

(c) Statistically, in terms of the everpresent power problem which, of course, is also relevant here, given that the sample sizes are so small.

(d) Statistically, on more technical grounds: The confidence bounds based on the t-distribution Jinks and Fulker report are invalid because they are derived on the assumption of the fixed effects ANOVA model, not the variance components model used by Jinks and Fulker. In the fixed effects model, the dependent variables are assumed to be normal, and the LSEs, as linear combinations of them, are also normal. Hence their variances are proportional to the diagonal elements of the inverse of  $X'X$ , (where  $X$  is the design matrix), and the ratios of LSEs over the square roots of the product of these diagonal elements multiplied by MSE are t. However, in the random effects model, the

TABLE 3. Parameter estimates (LSEs) under the genetic model JFG3 and the correlated environments model CE3.

		GENETIC MODEL				ENVIRONMENTAL MODEL			
		$\hat{\text{var}}(g)$	$\hat{\text{var}}(c_b)$	$\hat{\text{var}}(c_w)$	res	$\hat{\text{var}}(e)$	$\hat{\text{var}}(d)$	$\hat{\text{var}}(s)$	res
Neuroticism									
males	SS	11.89	-11.43	12.13	.09	.82	6.42	10.71	.00
females	SS	9.64	-2.55	9.94	.01	6.68	8.81	3.48	.00
pooled	SS	10.16	-4.27	10.26	.04	5.67	8.32	4.77	.03
pooled	MS	10.98	-7.41	11.36	.08	3.73	7.65	7.09	.04
J&F	MS	6.12	-2.06	9.02		(p. 329)			
Extraversion									
males	SS	9.54	-2.14	6.15	.00	8.19	5.08	.56	.01
females	SS	11.29	-7.16	10.21	.07	3.73	7.02	8.09	.00
pooled	SS	10.91	-6.18	9.45	.06	4.54	6.63	6.61	.02
pooled	MS	10.49	-4.78	8.47	.04	5.94	6.09	4.32	.03
J&F	MS	10.80	-2.47	8.31		(p. 333)			
Mill Hill									
males	SS	28.29	-19.62	19.63	.14	4.78	7.67	25.78	.01
females	SS	24.77	-9.48	10.87	.04	14.35	5.93	11.26	.00
pooled	SS	25.67	-11.67	12.66	.07	12.53	6.34	14.36	.02
pooled	MS	27.05	-16.00	14.36	.11	9.78	6.36	18.52	.03
J&F	MS	not given							
Dominoes									
males	SS	44.63	-.84	14.34	.01	38.34	13.82	9.50	.01
females	SS	80.69	-34.31	39.80	.02	47.14	21.92	32.83	.00
pooled	SS	71.95	-25.87	34.16	.04	45.84	20.11	26.27	.03
pooled	MS	63.20	-18.96	26.56	.06	43.14	17.08	21.17	.05
J&F	MS	59.61	-10.33	27.22		(p. 337)			

Notes: J&F = LSEs in Jinks and Fulker (1970) on page indicated. Pooled = pooled across both sexes. SS = sums of squares solution. MS = mean squares solution.  $\text{res} = 1 - \eta^2 = e'e/y'y$ .

TABLEAU 3. Estimations des paramètres sous le modèle génétique JFG3 et le modèle d'environnements corrélé, CE3.

dependent variables are sums of squares (or mean squares), and thus chi-square, not normal. The exact distribution of arbitrary linear combinations of such chi-square variates is probably rather difficult to establish (Satterthwaite, 1946), certainly it is not normal. Hence the confidence bounds Jinks and Fulker computed around their negative variance estimates are unsound statistically and any inferences based on them are invalid.

(e) In terms of prior odds: As was already the case for the sums and difference of the mean squares and the intraclass correlations, one is again struck by the consistency of these inadmissible variance estimates: the LSEs for  $\text{var}(e_b)$  are negative in all 8 of the smaller samples and all 4 pooled samples. While one expects negative variance estimates of a zero variance half the time, it is disturbing if such negative estimates arise in 12 out of 12 cases.

We now turn to the fit of the 3-parameter correlated environments model (CE3). The expected sums of squares are derived in Appendix A5 and the parameter estimates are summarized in Table 3. As can be seen, none of the parameter estimates are inadmissible. Reliability estimates for the four dependent variables are given by  $r_T$  under CE3 (see Appendix A5). Although Jinks and Fulker do not report the reliabilities of these variables, some of them can be inferred from other data based on comparable tests. For example, Osborne (1980) reported a wealth of data for black and white twins which include, among other tests, Cattell's so-called "Culture Fair Intelligence Test". This test is very similar to the Dominoes used by Shields because both are modelled after Raven's Progressive Matrices. For the female white sample ( $N = 66$ ), the parallel form reliability is .77, for the male white sample ( $N = 60$ ) it is .73. As can be seen from Table 4, the reliability estimated based on  $r_T$  are in the same range for the two cognitive variables. As one might expect, they are somewhat lower for the two self-rating questionnaire variables.

In Table 4, the fit of both models is compared in terms of the unweighted residuals ( $e'e/y'y = 1 - \eta^2$ , where now  $e$  is the least squares error vector and  $y$  the criterion vector) and also in terms of the predicted and observed intraclass correlations for the MZTs and MZAs. To predict them from the genetic model JFG3, negative variance estimates were replaced by zero, resulting in equal predictions for the MZTs and MZAs. As can be seen, regardless which measure of fit one may choose, the correlated environments model fits roughly twice as well as the genetic model proposed by Jinks and Fulker. In terms of the predicted intraclass correlations (left columns of Table 4), the ratio JFG3/CE3 of the sums of squared discrepancies is 2.17, and in terms of the error measures  $1 - \eta^2$  left by both models after a least squares fit (right columns of Table 4) it is 2.10 for the four larger, pooled groups. For the eight smaller within sex analyses it is 4.54.

This is not surprising, because we found early on that the Jinks and Fulker model consistently makes the wrong qualitative predictions. In particular, the inclusion of a separation variable  $s$  uniformly improves the differential prediction of the two intraclass correlations,  $r_A$ ,  $r_T$ , as it was designed to do, instead of predicting the same intermediate value for both, as does the Jinks and Fulker model for  $\text{var}(e_b) = 0$ .

TABLE 4. Comparison of fit between the genetic model JFG3 and the correlated environments model CE3.

Test Group		Intraclass correlations					LS fit of models		
sex	twin type	obs. r	predicted		difference		$e'e/y'y = 1 - \eta^2$		
			JFG3	CE3	JFG3	CE3	JFG3	CE3	ratio
Neuroticism									
male	together	.01	.50	.11	-.49	-.10	.09	.00	19.96
	apart	.71	.50	.64	.21	.07			
female	together	.46	.49	.43	-.03	.03	.01	.00	14.66
	apart	.50	.49	.54	.01	-.04			
pooled							.04	.03	1.73
Extraversion									
male	together	.53	.61	.62	-.08	-.08	.00	.01	.18
	apart	.73	.61	.63	.12	.10			
female	together	.38	.53	.35	-.15	.03	.07	.00	100.07
	apart	.60	.53	.63	.07	-.03			
pooled							.06	.02	3.27
Mill Hill									
male	together	.70	.60	.37	.10	.33	.14	.01	27.91
	apart	.73	.60	.81	.13	-.08			
female	together	.76	.70	.71	.06	.05	.04	.00	47.51
	apart	.78	.70	.81	.08	-.03			
pooled							.07	.02	4.31
Dominoes									
male	together	.90	.76	.74	.14	.16	.01	.01	2.45
	apart	.70	.76	.78	-.06	-.08			
female	together	.65	.67	.68	-.02	-.03	.02	.00	55.52
	apart	.81	.67	.78	.14	.03			
pooled							.04	.03	1.56
Sums of squared residuals:					.41	.19	.21	.10	(pooled only)
JFG3/CE3 ratio:					2.17		2.10		

TABLEAU 4. Comparaison des ajustements entre le modèle génétique JFG3 et le modèle des environnements corrélé, CE3.

It should be emphasized that the uniformly better fit of the correlated environments model to the Shields data does not, of course, settle the nature/nurture question one way or another. Rather, all we *can* say at this point is that

(a) The Shields data systematically violate critical qualitative predictions of the genetic model Jinks and Fulker fitted to them. As a result, they obtained an

inordinately high proportion of negative variance estimates. The confidence bounds they used to camouflage this defect are statistically unsound.

(b) In contrast, the correlated environments model produces consistently admissible parameter estimates and, moreover,

(c) it fits the Shields data quantitatively twice as well as the Jinks and Fulker model.

In conclusion, whatever the scientific import of the Shields data may be, they do not support the inferences Jinks and Fulker drew.

More recently, we found quite similar results for Osborne's (1980) personality test data of MZs and DZs (Schönemann & Schönemann, 1988; Schönemann, 1989). In this case, the fit of a purely environmental model surpassed that of the more traditional genetic model by a factor of 14. Not surprisingly, the most dominant variance component was measurement error.

## 6. DISCUSSION

The more general lesson is that purely descriptive variance component models, even if they *were* used properly with due regard to all assumptions, remain intrinsically inconclusive about heritability questions simply because one can never rule out that some other model may fit the data even better. Presumably, this is what Kempthorne (1990), a leading quantitative geneticist to whom Jinks and Fulker (1970) repeatedly appealed, had in mind when he cautioned "that most of the literature on heritability in species that cannot be experimentally manipulated, for example, in mating, should be ignored" (p. 139).

At the very least this literature should be treated with caution, especially since lately signs have multiplied that it cannot be trusted for other reasons: It now appears that patently absurd research claims were perpetuated for decades because valid evidence which conflicts with them had been suppressed. The surprising longevity of the flawed claims of Jinks and Fulker is just one case in point (see Hirsch, 1990, for more details) but not the only one. Another example is the recent dramatic revision of the 80% figure for IQ heritability which had been help up as a natural constant rivalling the speed of light for almost a century: "The heritability of the IQ -- that is, the percentage of individual differences variance attributable to genetic factors -- comes out to about 80 per cent, the average value obtained from all relevant studies now reported" (Jensen, 1970, testifying before the US Congress). We now learn from Henderson (1982) that "Today an estimate of 50% seems more in vogue" (p. 411) -- which puts faulty heritability claims on par with fashion trends. Another instance is the recent discovery that one of the most widely quoted heritability indices,



Holzinger's  $h^2$ , is mathematically unsound because Holzinger made a mistake which human behavior geneticists and statisticians alike had overlooked for over 50 years (Schönemann, 1988).

Thus it appears that some human behavior geneticists may not always have realized the full scope of potentially adverse social implications that their unwarranted inferences may have, and, as the record shows, already have had (Blum, 1978; Chase, 1980; Robitscher, 1973; Shields, 1978).

#### ACKNOWLEDGMENTS

This paper had an instructive history which will be described elsewhere. I wish to thank, first and foremost, Professor Jerry Hirsch, University of Illinois, for his constant encouragement and advice, and, more generally, for his inspiring example of socially responsible scholarship. I also would like to thank my wife Roberta and two unknown reviewers for their careful reading of earlier drafts and their help in weeding out technical ambiguities. Finally, I would like to take this opportunity to thank Professor Robert Brush, Purdue University, for representing me at the recent Behavior Genetics Association Meetings.

#### RESUME

*On montre que le modèle génétique de Jinks et Fulker (1970), adapté aux données de Shields (1962) sur des jumeaux, est qualitativement incompatible avec des tendances systématiques observées dans ces données ; en conséquence, ce modèle produit une proportion inhabituelle d'estimations de variance négatives. Par opposition, un modèle purement environnemental produit des prédictions qualitatives compatibles avec les données de Shields, ainsi que des estimations de paramètres généralement admissibles. Quantitativement, ce modèle s'ajuste beaucoup mieux aux données de Shields que ne le fait le modèle génétique de Jinks et Fulker. Dans l'état actuel, les conclusions de ces auteurs, d'une portée considérable, ne sont pas prouvées par les données de Shields. Cette réévaluation illustre le fait que des modèles purement descriptifs, même utilisés avec circonspection, ne permettent pas de se prononcer définitivement sur l'opposition entre nature et culture, car il reste possible que d'autres modèles puissent s'ajuster à ces mêmes données.*

#### REFERENCES

- Blum, J.F. (1978). *The pseudoscience of mental ability*. New York: Monthly Review Press.
- Chase, A. (1980). *The legacy of Malthus*. Urbana: University of Illinois Press.
- Eysenck, H.J. (1973). *The measurement of intelligence*. Baltimore: Williams and Wilkins.
- Farber, S.L. (1981). *Identical twins reared apart. A reanalysis*. New York: Basic Books.

- Fisher, R.A. (1921). On the "probable error" of a correlation deduced from a small sample. *Metron*, 1, 3-32.
- Galton, F. (1883). *Inquiries into human faculty and its development*. New York: McMillan.
- Harris, J.A. (1913). On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika*, 9, 446-472.
- Henderson, N.D. (1982). Human behavior genetics. *Annual Review of Psychology*, 33, 403-440.
- Hirsch, J. (1981). To "unfrocks the charlatans". *Sage Race Relations Abstracts*, 6, 1-67.
- Hirsch, J. (1990). A nemesis for heritability estimation. *Behavioral and Brain Sciences*, 13, 137-138.
- Jensen, A.R. (1970). Statement of Dr. Arthur A. Jensen before the General Subcommittee on Education, House Education and Labor Committee. Washington, *Congressional Record*, I-237, H 6319, H 6325, H 6326.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- 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.
- Kamin, L.J. (1974). *The science and politics of IQ*. New York: Erlbaum.
- Kemphorne, O. (1990). How does one apply statistical analysis to our understanding of the development of human relationships. *Behavioral and Brain Sciences*, 13, 138-139.
- King, J.C. (1981). *The biology of race*. Berkeley: University of California Press.
- Loehlin, J.C., Lindzey, G., & Spuhler, J.N. (1975). *Race differences in intelligence*. San Francisco: Freeman.
- Morgan, T.H. (1929). The mechanisms and laws of heredity. In C. Murchison (Ed.), *Foundations of experimental psychology* (pp. 1-44). Worcester: Clark University Press.
- Osborne, R.T. (1980). *Twins: Black and white*. Athens: Foundation for Human Understanding.
- Robitscher, J. (1973). *Eugenic sterilization*. Springfield: Thomas.
- Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2, 110-114.
- Scheffe, H.A. (1959). *The analysis of variance*. New York: Wiley.
- Schönemann, P.H. (1987). On non-null tests of intraclass correlations. Purdue Mathematical Psychology Program Technical Report N° 87-6. Purdue University, West Lafayette, Indiana.
- Schönemann, P.H. (1988). A note on Holzinger's heritability estimate  $h^2$ . Purdue Mathematical Psychology Program Technical Report N° 88-5. Purdue University, West Lafayette, Indiana.
- Schönemann, P.H. (1989). New questions about old heritability estimates. *Bulletin of the Psychonomic Society*, 27, 175-178.
- Schönemann, P.H. (in press). In praise of randomness. *Behavioral and Brain Sciences*.
- Schönemann, P.H., & Schönemann, R.D. (1988). Environmental versus genetic models of Osborne's personality data on identical and fraternal twins. Purdue Mathematical Psychology Program Technical Report N° 88-6. Purdue University, West Lafayette, Indiana.

- Shields, J. (1962). *Monozygotic twins brought up apart and brought up together*. London: Oxford University Press.
- Shields, J. (1978). MZA twins: Their use and abuse. In W. Nance (Ed.), *Twin research psychology and methodology*. New York: Liss.
- Snedecor, G.W., & Cochran, W.G. (1967). *Statistical methods*. Ames, Iowa: The Iowa State University Press, 6th edition.
- Wahlsten, D. (1990). Insensitivity of the analysis of variance to heredity-environment interaction. *Behavioral and Brain Sciences*, 13, 109-161.
- Winer, B.J. (1971). *Statistical principles of experimental design*. New York: McGraw Hill.

Received September 25, 1989

Accepted August 28, 1990

## APPENDICES

### A1: SUMS OF SQUARES OF $N \times 2$ TWIN SCORE MATRICES

Let  $y_{ik}$  be the observed score of the  $k$ 'th twin ( $k = 1, 2$ ) in the  $i$ 'th family ( $i = 1, N$ ), and  $X = (x_{ij})$  be the  $N \times 2$  matrix of deviation scores  $x_{ik} := y_{ik} - \bar{y}$  around their grand mean  $\bar{y}$ . On denoting the two columns of the deviation score matrix  $X$  by  $x_1, x_2$ , one has the identity

$$x_1'x_1 + x_2'x_2 = (x_1 + x_2)'(x_1 + x_2)/2 + (x_1 - x_2)'(x_1 - x_2)/2.$$

By definition, the total sum of squares is given by

$$SST := \sum_k \sum_i (y_{ik} - \bar{y})^2 = \sum_k \sum_i x_{ik}^2 = x_1'x_1 + x_2'x_2.$$

For the sum of squares between rows (twin pairs) one finds

$$\begin{aligned} SSB &:= \sum_k \sum_i (y_i - \bar{y})^2 = 2 \sum_i (y_i - \bar{y})^2 \\ &= 2 \sum_i [y_{i1} + y_{i2}]/2 - \bar{y}]^2 = \sum_i (y_{i1} + y_{i2} - 2\bar{y})^2/2 \\ &= (x_1 + x_2)'(x_1 + x_2)/2. \end{aligned}$$

By subtraction,

$$SSW = SST - SSB = (x_1 - x_2)'(x_1 - x_2)/2.$$

Thus,  $2SSB$  and  $2SSW$  are simply the sums of squares of the row sums and row differences of the  $N \times 2$  matrix of deviation scores around the grand mean  $\bar{y}$ . If all  $y_{ik}$  have the same mean  $\mu$ ,  $E(X) = \emptyset$ . Since  $SST$  is a quadratic form of rank  $2N-1$ , and  $SSW$  a quadratic form of rank  $N-1$ ,  $SSB$  has  $N-1$  df and  $SSW$   $N$  df.

### A2: INTRACLAS CORRELATIONS

Intraclass correlations are measures of resemblance between twins within pairs. In the sample, they can be defined

- (a) as the product moment correlation between the two columns of the  $2N \times 2$  matrix

$$X^* := \begin{pmatrix} x_1 & x_2 \\ x_2 & x_1 \end{pmatrix}$$

(Harris, 1913; Fisher, 1921; Snedecor & Cochran, 1967), so as to remove the ambiguity whether a twin's score should be assigned to the left or right column ("double entry table correlation"). In this case the intraclass correlation is given by

$$r' = (SSB - SSW)/(SSB + SSW).$$

Alternatively, (b), the intraclass correlation can be defined as

$$r = (MSB - MSW)/(MSB + MSW),$$

which is the definition usually adopted in twin research and used throughout in the present paper. One obtains a definition for the population by replacing the mean squares by their expected values. In Schönemann (1987) it is shown that the non-null distribution of  $s(\rho)/s(r) = s(\rho)MSB/MSW$ , where  $s(x) := (1-x)/(1+x)$ , is central  $F_{N-1, N}$ .

### A3: THE JINKS AND FULKER 4-PARAMETER GENETIC MODEL (JFG4)

Jinks and Fulker (1970) fit the variance component model

$$y - \mu := g_b + g_w + e_b + e_w$$

to all 3 groups of twins, MZTs, MZAs, and DZTs jointly. This model postulates for the

$$\text{MZTs: } y_1 - \mu = g_b + g_w + e_b + e_{w1} \quad (1\text{st twin})$$

$$y_2 - \mu = g_b + g_w + e_b + e_{w2} \quad (2\text{nd twin})$$

where all variables on the right have zero means,  $\text{var}(e_{wk}) = \text{var}(e_w)$ ,  $\text{cov}(g_b, g_w) = \text{cov}(g_b, e_k) = \text{cov}(g_w, e_k) = \text{cov}(e_1, e_2) = 0$ .

Hence

$$y_1 + y_2 - 2\mu = 2g_b + 2g_w + 2e_b + e_{w1} + e_{w2}$$

$$y_1 - y_2 = e_{w1} - e_{w2}$$

Together with the stated covariance assumptions, this implies for the expected sums of squares

$$E(SSBT) = (N_T - 1)[2\text{var}(g_b) + 2\text{var}(g_w) + 2\text{var}(e_b) + \text{var}(e_w)]$$

$$E(SSWT) = N_T \text{var}(e_w)$$

where  $N_T$  is the number of pairs of twins raised together, and SSBT has  $N_T - 1$  df and SSWT  $N_T$  df by A1.

$$\text{MZAs: } y_1 - \mu = g_b + g_w + e_{b1} + e_{w1} \quad (1\text{st twin})$$

$$y_2 - \mu = g_b + g_w + e_{b2} + e_{w2} \quad (2\text{nd twin})$$

with the same expected values and variance - covariance assumptions as for the MZTs, except that now also  $E(e_{bk}) = 0$  and the two  $e_{bk}$  are uncorrelated with each other and all other variables on the right. Hence

$$y_1 + y_2 - 2\mu = 2g_b + 2g_w + e_{b1} + e_{b2} + e_{w1} + e_{w2}$$

$$y_1 - y_2 = e_{b1} - e_{b2} + e_{w1} - e_{w2}$$

Together with the stated covariance assumptions, this model implies the expected sums of squares

$$E(SSBA) = (N_A - 1)[2\text{var}(g_b) + 2\text{var}(g_w) + \text{var}(e_b) + \text{var}(e_w)]$$

$$E(SSWA) = N_A[\text{var}(e_b) + \text{var}(e_w)],$$

where  $N_A$  is the number of twin pairs raised apart, and SSBA has  $N_A - 1$  df and SSWA  $N_A$  df by A1.

This model does not imply that the total variance (= total mean squares) are the same for MZT's and MZA's, as Jinks and Fulker claim (1970, p. 326). Rather,

$$E(\text{MSTT}) := E(\text{SSBT} + \text{SSWT})/(2N_T-1) \\ = (2N_T-2)[\text{var}(g_b) + \text{var}(g_w) + \text{var}(e_b)]/(2N_T-1) + \text{var}(e_w)$$

while

$$E(\text{MSTA}) := E(\text{SSBA} + \text{SSWA})/(2N_A-1) \\ = (2N_A-2)[\text{var}(g_b) + \text{var}(g_w)]/(2N_A-1) + \text{var}(e_b) + \text{var}(e_w),$$

so that  $E(\text{MSTT}) \neq E(\text{MSTA})$  in general, though the difference will be small since  $(2N-2)/(2N-1)$  is close to 1.

Instead, the model predicts that the two expressions based on mean squares,

$$E(\text{MSBT} + \text{MSWT}) := E(\text{MSBA} + \text{MSWA}) \\ = 2[\text{var}(g_b) + \text{var}(g_w) + \text{var}(e_b) + \text{var}(e_w)]$$

will be equal. It also predicts  $E(\text{MSBT}-\text{MSWT}) \geq E(\text{MSBA}-\text{MSWA})$ :

$$E(\text{MSBT} - \text{MSWT}) = 2[\text{var}(g_b) + \text{var}(g_w) + \text{var}(e_b)] \\ \geq 2[\text{var}(g_b) + \text{var}(g_w)] = E(\text{MSBA} - \text{MSWA}).$$

Consequently,  $\rho_T \geq \rho_A$ :

$$\rho_T := [\text{var}(g_b) + \text{var}(g_w) + \text{var}(e_b)]/[\text{var}(g_b) + \text{var}(g_w) + \text{var}(e_b) + \text{var}(e_w)] \\ \geq [\text{var}(g_b) + \text{var}(g_w)]/[\text{var}(g_b) + \text{var}(g_w) + \text{var}(e_b) + \text{var}(e_w)] =: \rho_A.$$

The last three predictions are consistently violated by the Shields data.

#### A4: THE 3-PARAMETER GENETIC MODEL (JFG3)

Since the 4 variance parameters of the genetic model JFG4 cannot be estimated on the basis of identical twin data alone, Jinks and Fulker fit the 3-parameter model

$$y - \mu = g + e_b + e_w$$

to the MZTs, which results from JFG4 on defining  $g := g_b + g_w$ . For this model, the expected sums of squares reduce to:

$$E(\text{SSBT}) = 2(N_T-1)[\text{var}(g) + \text{var}(e_b)] + (N_T-1)\text{var}(e_w)$$

$$E(\text{SSWT}) = N_T \text{var}(e_w)$$

$$E(\text{SSBA}) = 2(N_A-1)\text{var}(g) + (N_A-1)[\text{var}(e_b) + \text{var}(e_w)]$$

$$E(\text{SSWA}) = N_A [\text{var}(e_b) + \text{var}(e_w)].$$

Hence the least squares equations for estimating  $\text{var}(g)$ ,  $\text{var}(e_b)$ , and  $\text{var}(e_w)$  are

$$\begin{bmatrix} \text{SSBT} \\ \text{SSWT} \\ \text{SSBA} \\ \text{SSWA} \end{bmatrix} = \begin{bmatrix} 2N_T-2 & 2N_T-2 & N_T-1 \\ 0 & 0 & N_T \\ 2N_A-2 & N_A-1 & N_A-1 \\ 0 & N_A & N_A \end{bmatrix} \begin{bmatrix} \text{var}(g) \\ \text{var}(e_b) \\ \text{var}(e_w) \end{bmatrix} + \text{error}$$

Since the coefficient matrix contains a  $3 \times 3$  non-singular triangular submatrix, it has full column rank 3 for all  $N_A$ ,  $N_T$ , so that the least squares estimates (LSEs) are unique.

For the expected mean squares between twin types one finds:

$$E(MSTT) = E(MSTA), \text{ in general,}$$

$$E(MSBT + MSWT) = 2[\text{var}(g) + \text{var}(e_b) + \text{var}(e_w)] = E(MSBA + MSWA)$$

$$E(MSBT - MSWT) = 2[\text{var}(g) + \text{var}(e_b)] \geq 2\text{var}(g) = E(MSBA - MSWA)$$

$$\rho_T := [\text{var}(g) + \text{var}(e_b)] / [\text{var}(g) + \text{var}(e_b) + \text{var}(e_w)]$$

$$\geq \text{var}(g) / [\text{var}(g) + \text{var}(e_b) + \text{var}(e_w)] =: \rho_A$$

Note that  $\text{var}(e_b) = 0$  implies  $\rho_T = \text{var}(g) / [\text{var}(g) + \text{var}(e_w)] = \rho_A$ . The last four predictions are consistently violated by the Shields data.

#### A5: THE 3-PARAMETER CORRELATED ENVIRONMENTS MODEL (CE3)

$$y - \mu = e + d + s$$

postulates an *environmental* variable  $e$ , a measurement error variable  $d$ , and a separation variable  $s$  which affects only the MZAs (separated twins). If one further assumes that the two environment variables  $e_1, e_2$  are perfectly correlated for the MZAs, one finds, for the

$$\text{MZTs: } y_1 - \mu = e + d_1 \quad (1\text{st twin})$$

$$y_2 - \mu = e + d_2 \quad (2\text{nd twin})$$

where the variables on the right of each equation have zero means,  $\text{var}(d_1) = \text{var}(d_2)$ ,  $\text{cov}(e, d_k) = \text{cov}(d_1, d_2) = 0$ . Hence

$$y_1 + y_2 - 2\mu = 2e + d_1 + d_2$$

$$y_1 - y_2 = d_1 - d_2.$$

Hence the expected sums of squares are

$$E(SSBT) = (N_T - 1)[2\text{var}(e) + \text{var}(d)]$$

$$E(SSWT) = N_T \text{var}(d).$$

$$\text{MZAs: } y_1 - \mu = e_1 + d_1 + s$$

$$y_2 - \mu = e_2 + d_2 + s$$

where the three variables on the right of each equation have zero means,  $\text{var}(e_1) = \text{var}(e_2) = \text{var}(e)$ ,  $\text{var}(d_1) = \text{var}(d_2) = \text{var}(d)$ ,  $\text{cov}(s, d_k) = \text{cov}(s, e_k) = \text{cov}(e_k, d_j) = \text{cov}(d_1, d_2) = 0$ , but  $\text{cov}(e_1, e_2) = \text{var}(e)$ , so that the two environmental variables affecting two separated twins are perfectly correlated (in contrast, the genetic models JFG4 and JFG3 assume that  $e_1$  is perfectly *uncorrelated* with  $e_2$ ). Hence

$$y_1 + y_2 - 2\mu = e_1 + e_2 + d_1 + d_2 + 2s$$

$$y_1 - y_2 = e_1 - e_2 + d_1 - d_2.$$

This implies the expected sums of squares

$$E(SSBA) = (N_A - 1)[2\text{var}(e) + \text{var}(d) + 2\text{var}(s)]$$

$$E(SSWA) = N_A \text{var}(d),$$

since  $\text{var}(e_1 + e_2) = 4 \text{var}(e)$  for MSBA, while for MSWA the covariance term cancels against  $\text{var}(e)$ .

Hence the expected mean squares between twin types are

$$E(MSBT + MSWT) = 2[\text{var}(e) + \text{var}(d)]$$

$$\leq 2[\text{var}(e) + \text{var}(d) + \text{var}(s)] = E(MSBA - MSWA)$$

$$E(MSBT - MSWT) = 2\text{var}(e)$$

$$\leq 2[\text{var}(e) + \text{var}(s)] = E(MSBA - MSWA).$$

The intraclass correlations are

$$\rho_T = \text{var}(e)/[\text{var}(e) + \text{var}(d)] \\ \leq [\text{var}(e) + \text{var}(s)]/[\text{var}(e) + \text{var}(d) + \text{var}(s)] = \rho_A.$$

These ordinal predictions agree with the Shields data. Note that  $\rho_T$  is simply the reliability of  $y$ .

The least squares equations for this model are

$$\begin{bmatrix} \text{SSBT} \\ \text{SSWT} \\ \text{SSBA} \\ \text{SSWA} \end{bmatrix} = \begin{bmatrix} 2N_T-2 & 2N_T-1 & 0 \\ 0 & N_T & 0 \\ 2N_A-2 & N_A-1 & 2N_A-2 \\ 0 & N_A & 0 \end{bmatrix} \begin{bmatrix} \text{var}(e) \\ \text{var}(d) \\ \text{var}(s) \end{bmatrix} + \text{error}$$

The coefficient matrix has always full column rank 3 for the same reasons as that of JFG3 so that the LSEs are unique. If the models are fitted to the males and females jointly, the coefficient matrices are replicated twice for JFG3 and CE3.