CPC, 1994, 13 (2), 141-167

Environmental versus genetic models for Osborne's personality data on identical and fraternal twins

PETER H. SCHÖNEMANN[1]
and ROBERTA D. SCHÖNEMANN[2]

1. Department of Psychological Sciences, Purdue University
2. Departments of Mathematics and Science Counseling, Purdue University

Abstract. *It is shown that the additive genetic model of Nichols needed to justify the heritability ratio HR does not fit Osborne's (1980) personality data very well. A purely environmental model with the same number of parameters fits these data better by a factor of 14. Compared with the additive genetic model, these empirical results suggest that Osborne's personality data contain no genetic component at all. The responses of identical twins may be more similar simply because they are exposed to more similar environments than fraternal twins. This outcome illustrates the general principle that conventional variance component models used to justify heritability estimates are intrinsically inconclusive: We can never rule out that another, qualitatively quite different model fits the same data equally well or, as in the present case, much better.*

Key words: Cognition, heritability, genetics, twin studies, personality.

Correspondence should be sent to Peter H. Schönemann, Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, U.S.A. (e-mail: phs @ psych.purdue.edu).

1. INTRODUCTION

In 1970 Jensen informed the General Subcommittee on Education, House Education and Labor Committee that "individual differences in intelligence, that is, IQ, are predominantly attributable to genetic differences, with environmental factors contributing a minor portion of the variance among individuals. The heritability of the IQ ... comes out to about 80 percent, the average value obtained from all relevant studies now reported" (Jensen, 1970, p. H6325).

This assessment was shared by well known population geneticists: "The twin data, however, leave no serious doubt that heritability [of IQ] was at least 70% and perhaps as great as 80% in the populations studied ... Similarly high total heritabilities have been derived by most authors, irrespective of the precise formula used for h^2" (Wright, 1978, p. 414). As Wright noted, most of these estimates were derived from twin studies. Broadly, these studies fall into two main categories: (a) those comparing monozygotic twins raised together (MZTs) with monozygotic twins raised apart (MZAs), and (b) those comparing MZTs with dizygotic twins raised together (DZTs).

The empirical basis of twin studies of type (a) is exceedingly small. It shrank further when Kamin discovered that more than half of the then extant MZA data, those reported by Burt, were untrustworthy (Kamin, 1974; Hearnshaw, 1981). The largest remaining MZA study, by Shields (1962), contains only 44 pairs of MZAs. A recent re-examination of Shields' data in (Schönemann, 1990) showed that the conventional variance component model which underlies the heritability ratios (e.g., Jinks & Fulker, 1970) is systematically violated by these data:

(a) Contrary to the predictions of this model, twins raised in the same environment resembled each other less, not more, than twins raised in different environments, and

(b) the estimates of the environmental variance components turned out negative in 12 out of 12 subsets of the data.

(c) On fitting the same data to a purely environmental model which postulates no genetic components at all, the fit improved by a factor of 2.

The main problem with MZT/MZA data, apart from the necessarily small sample sizes for MZAs, is that twins raised apart are not randomly assigned to different environments. Once this assumption is violated, the underlying variance component model, and thus the heritability ratios derived from it, become invalid.

In the present paper, we turn our attention to the second type of twin studies, those comparing monozygotic and dizygotic twins raised together (MZTs and DZTs, MZs and DZs from now on). Such data are more interesting than those of MZTs and MZAs because

(a) the sample sizes are usually much larger, and

(b) MZ/DZ data agree at least with one basic prediction of the classical random components model: MZs should resemble each other more than DZs, since MZs, by definition, share exactly the same genes, while DZs share only half the genes on average.

Twin resemblance is usually measured in terms of a so-called "intraclass correlation coefficient" which, for twins, varies between -1 and 1. Its sample estimate can be computed in terms of mean squares as

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

where MSB denotes the mean square between (families) and MSW the mean square within (see Appendix 1 for details). An intraclass correlation can be viewed as an ordinary correlation computed from a "double entry table" which contains each twin pair twice, once as (y_{i1}, y_{i2}) and once as (y_{i2}, y_{i1}), so as to remove the ambiguity as to whether a twin's score should appear on the left or on the right. As Osborne (1980, p. 16) notes, one usually finds that the intraclass correlation r_M for MZs raised together exceeds the correlation r_D for DZs raised together by roughly .2 over a wide range of cognitive and personality tests.

If certain additivity assumptions are met by the data, then the sample intraclass correlations r_M, r_D can be used to estimate "broad heritability",

$$h_B := var(g)/[var(g)+var(e)],$$

i.e., the proportion the genetic component g contributes to the variance of the observed score $y := g + e + z$, where e is the environmental component, with the measurement error component z removed. Nichols (1965) has proposed the statistic

$$HR := 2(r_M - r_D)/r_M$$

as an estimate of h_B. As shown in Appendix 3, HR follows as a valid deduction from a strictly additive genetic variance component model. In contrast, the older coefficient,

$$h^2 := (r_M - r_D)/(1 - r_D),$$

proposed by Holzinger (Newman, Freeman, & Holzinger, 1937), is not a valid heritability estimate, because Holzinger's derivations were unsound. As was shown in more detail in Schönemann (1989), his derivations imply that dizygotic twins share no genes. This conflicts with standard genetic theory that they share half the genes. Consequently, we shall focus on HR in this paper.

In terms of HR, the MZ/DZ data raise some perplexing questions:

(a) The HRs tend to be much higher for mental tests than for biological variables used in controlled animal breeding experiments. King (1981) reports heritability estimates for milk production in cattle of the order of .3, those of egg production in poultry rarely exceed .5. In contrast, Jensen (above) reports an overall average heritability of .8 for "intelligence". For personality tests, one often finds that HR exceeds 1.

(b) The magnitude of HR tends to increase inversely with the quality of the tests: Osborne (p. 22, loc. cit.) reports intraclass correlations implying HRs of .63 for "General Ability" measured by the National Merit Scholarship Qualification Test. The averages for the presumably less reliable personality tests are .98. Those for "Goals and Ideals" and for "Self Concept" rating scales even exceed 1. Rowe (1981, p. 205) reports MZ/DZ data for the "Perceived Parental Dimension, Perception of Father and Perception of Mother" implying HRs of 1.43 and 1.37. For Cattell's High School Personality Questionnaire (Osborne, 1980, p. 127), 23 out of 42 (= 54.8%) heritability ratios exceed 1. Altogether 69% of the ratios are inadmissible (i.e., outside the range [0,1]). For Jenkins' Personality Test, (1959) "How Well Do You Know Yourself", this proportion rises to 76%. The majority of the estimates are again too large.

(c) The heritability ratios HR are often high even for variables for which it is difficult to imagine any genetic components. For example, Loehlin and Nichols (1976) asked large samples of male and female MZs and DZs to indicate "things you have done during the past year" (Appendices, p. 6) and reported the intraclass correlations. For the item "Baby sat" (item 102), one finds HR = .35 for the males and .47 for the females. For item 239, "Used a thermometer and took your temperature," the two HRs were .89 and .88. For item 250, "Had your back rubbed," the HRs were .92 and .21. For all these items, the intraclass correlations were higher for the MZs.

2. THE STRICTLY ADDITIVE GENETIC MODEL AG3

As shown in Appendix A3, Nichols' (1965) heritability ratio HR defined by

$$HR := 2(r_M - r_D)/r_M$$

can be justified as an estimate of "narrow heritability" if a 3-parameter variance component model fits the data, which postulates three types of latent variables: genetic variables a_k, an environmental variable e, and

measurement error variables z_k. This is an adaptation to MZ/DZ data of the genetic model Jinks and Fulker (1970) used to describe the Shields (1962) MZT/MZA data, although with dubious success (Schönemann, 1990).

This model (see Fig. 1A) explains the observed scores y_1, y_2 of MZs by

$$\text{MZs:}\quad y_1 - \mu = a + e + z_1, \qquad y_2 - \mu = a + e + z_2$$

under the assumption that the genetic, environmental, and measurement error variables have zero means and are uncorrelated (see left side of Fig. 1A). The observed scores y_1, y_2 of two DZs are explained by

$$\text{DZs:}\quad y_1 - \mu = a_1 + e + z_1, \qquad y_2 - \mu = a_2 + e + z_2$$

under the added assumption that the correlation between a_1 and a_2 is 1/2 in accordance with standard genetic theory (right side of Fig. 1A). Since this model includes genetic variables and fits three parameters, we denote it "AG3".

The model AG3 implies a number of ordinal predictions about the mean squares and intraclass correlations which should be checked before one attaches meaning to HR (see Appendix 3):

(a) E(MSBM + MSWM) = E(MSBD + MSWD),

(b) E(MSBD - MSWD) ≥ E(MSBM - MSWM)/2.

Conditions (a), (b) together imply the stronger condition

(c) $0 \le \rho_M/2 \le \rho_D \le \rho_M \le 1$,

where ρ_M denotes the population intraclass correlation for the MZs, and ρ_D for the DZs.

The model AG3 also implies the identity

$$\text{var(a)}/[\text{var(a)}+\text{var(e)}] = \text{HR},$$

i.e., Nichol's heritability ratio HR coincides with "narrow heritability" h_A if (and only if) AG3 fits the data.

Jensen (1967, p. 149) dismisses both Holzinger's h^2 (which he denotes "H") and HR with the curious logic "that one is not a monotonic function of the other". But the fact that two answers to the same question disagree (e.g., the answers "2" and "3" to 1+1 = ?) does not rule out that one of them might be correct. Nichols' HR is an estimate of narrow heritability, if and only if, AG3 fits the data, regardless how it relates to other, incorrect heritability estimates.

Jensen has modified Nichols' HR to arrive at yet a third heritability estimate, which we denote

$$JHR := 2(r_M - r_D),$$

but without supplying an explicit variance component model from which it could be derived. JHR could be viewed as a limiting case of HR when $r_M = 1$ (in which case, it should be written $2(1-r_D)$). However, to

impose such a condition on Nichols' HR would further restrict its prospects for empirical verification, which are already dim enough without imposing additional constraints.

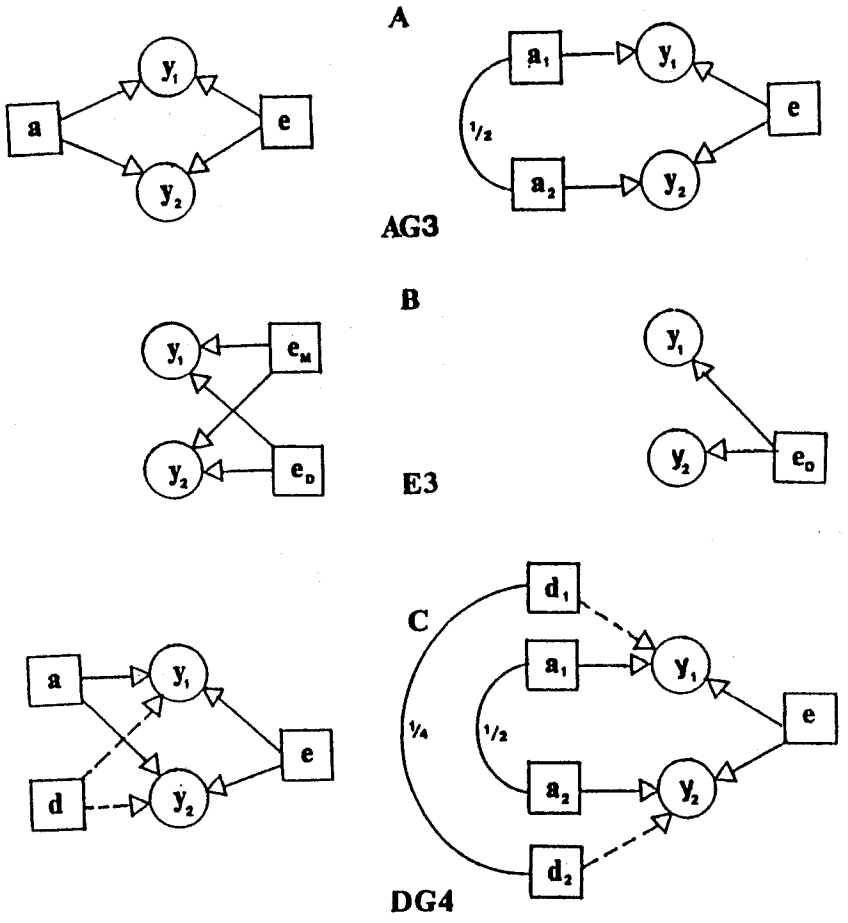

Figure 1. Three models for identical and fraternal twins. A: The additive genetic model AG3. B: the purely environmental model E3. C: Fisher's dominance model with measurement error, DG4.

Left side: MZs. Right side: DZs. Boxes = independent variables (Genetic variables: a = additive values, d = dominance deviations. The measurement errors z_k have been omitted for clarity. Environmental variables: e, e_D = common environment, e_M = additional common environment induced by physical resemblance of MZs). Curved lines = correlations between independent variables. Circles = dependent variables: y_1, y_2 = scores of both twins.

3. THE 3-PARAMETER ENVIRONMENTAL VARIANCE COMPONENTS MODEL E3

The fact that the intraclass correlations of MZs generally exceed those of DZs agrees with the hereditarian premise that mental tests contain a substantial genetic component. However, as several authors already noticed in the 30s, this is by no means the only possible explanation of the greater resemblance of MZs. For example, Jones (1949) proposed an alternative explanation:

"... we must also take into account the fact of a greater degree of environmental similarity. Several studies have shown that identical twins spend more time together, enjoy more similar reputations, are more likely to be in the same classrooms, have more similar health records, and, in many other respects, share a more common physical and social environment than ordinarily experienced by fraternal twins (Stocks, 1930-1931; Jones and Wilson, 1932-1933; Wilson, 1934; Lehtovaara, 1938)" (Jones, 1949, p. 613).

The main purpose of the present paper is to assess the relative plausibility of these two contrasting explanations of MZ/DZ twin data. To capture the essence of the environmental hypothesis, we postulate a 3-parameter environmental variance component model (E3) which contains no genetic variables at all, and which has exactly the same number of parameters as the genetic model AG3, so that the relative fit of both models can be compared:

To predict larger intraclass correlations for MZs than for DZs, we postulate two environmental variables, a baseline environmental variable e_D, which enters the scores of both MZs and DZs, and a second component e_M, which only enters the scores of the MZs, thus raising their environmental variance. In addition, we again postulate measurement error variables z_k to account for the differences in reliability of the various personality tests. Specifically, we postulate for the MZs:

$$\text{MZs:}\quad y_1 - \mu = e_M + e_D + z_1, \qquad y_2 - \mu = e_M + e_D + z_2,$$

under the stochastic restriction that all variables on the right of the equality signs are pairwise uncorrelated and have zero means (left side of Fig. 1B). The model for the DZs is:

$$\text{DZs:}\quad y_1 - \mu = e_D + z_1, \qquad y_2 - \mu = e_D + z_2,$$

(right side of Fig. 1B). As shown in Appendix 4, this implies

$$E(MSBM+MSWM) \geq E(MSBD+MSWD)$$

$$E(MSBM-MSBD) = 2var(e_M) \geq 0,\ E(MSWM) = var(z) = E(MSWD),$$

$$1 \geq \rho_M = [var(e_M)+var(e_D)] / [var(e_M)+var(e_D)+var(z)]$$

$$\geq var(e_D)/ [var(e_D)+var(z)] = \rho_D \geq 0.$$

4. RELATIVE FIT OF AG3 VERSUS E3 FOR OSBORNE'S PERSONALITY DATA

Osborne (1980) has provided a wealth of physical, cognitive, and personality data for same-sex monozygotic and dizygotic twins. In this report, we focus on his personality test data, which are conveniently summarized in Tables XII-E and XII-I (p. 127 and p. 135). Our interest in this particular subset of the Osborne data was aroused by the excessively large proportion of inadmissible heritability ratios HR.

Table XII-E (p. 128, loc. cit.) gives the results for relatively large samples of teenage male and female MZs and DZs for Jenkins' personality questionnaire "How Well Do You Know Yourself?" (HWDYKY, Jenkins, 1959). Osborne reports the intraclass correlations for its 17 subscales and 2 "control scales" (p. 126), the sample sizes, the heritability ratios h^2 and HR, and the mean squares within both twin types. This information permits reconstruction of the between sums of squares within the accuracy of the reported figures (2 decimal places), so that various variance component models can be fitted to these data, which are summarized in Table 1A. Columns 1 and 2 give the intraclass correlations and columns 3-6 the mean squares. For the content of the 19 subscales, see loc. cit.

In Table XII-I (p. 135, loc. cit.), Osborne gives the results for teenage white and black MZs and DZs and for the pooled sample for Cattell's "High School Personality Questionnaire" (HSPQ, Cattell, 1958, 1969). This is a junior version of his well-known "16PF" designed to "measure" 14 personality factors. The reported intraclass correlations, sample sizes, and within mean squares, and the recomputed between mean squares are given in Table 1B (Columns 1-6). Of particular interest is "Factor B" (test number 2 in Table 1B), because it is an abbreviated, unspeeded "Intelligence Test".

Altogether we deal with 5 data sets: two for the Jenkins test (denoted JM for males, JF for females), and three for Cattell's HSPQ (HW for the white sample, HB for the black sample, and HT for the total, pooled sample). The most striking aspect of all of them is the large proportion of inadmissible heritability ratios HR, which are summarized in Table 2. Overall, only a fourth of the values lie between 0 and 1, the majority (60%) are larger than 1.

TABLE 1. A: Jenkins' personality questionnaire.

MALES (JM, N_M = 78, N_D = 48)

	Correlations		Mean squares				Parameter estimates						
							AG3			E3			
	r_M	r_D	MSBM	MSWM	MSBD	MSWD	v(a)	v(e)	v(z)	$v(e_M)$	$v(e_D)$	v(z)
1	.37	-.10	30.25	13.91	15.58	19.04	19.80	-11.63	11.52	7.34	-.45	16.48
2	.40	.14	37.03	15.87	14.32	10.80	17.64	-7.06	8.93	11.35	.49	13.34
3	.07	-.02	29.57	25.70	23.60	24.56	4.83	-2.89	23.92	2.98	-.76	25.13
4	.20	.14	26.67	17.78	36.12	27.25	.02	4.42	22.51	-4.72	6.80	22.52
5	.32	-.02	42.08	21.68	18.78	19.55	21.17	-10.97	15.32	11.65	-.92	20.61
6	.47	.18	36.83	13.28	21.69	15.07	16.93	-5.16	9.94	7.57	3.76	14.18
7	.32	-.07	40.92	21.08	19.19	22.08	22.73	-12.81	15.90	10.86	-1.19	21.58
8	.02	.01	20.86	20.04	21.44	21.02	.40	.01	20.43	-.29	.45	20.53
9	.27	-.03	24.41	14.03	16.28	17.29	11.39	-6.20	12.81	4.06	.31	15.66
10	.25	.20	28.38	17.03	22.62	15.08	3.81	1.87	15.10	2.88	3.28	16.06
11	.47	.04	34.25	12.35	13.76	12.70	20.84	-9.89	7.32	10.24	.62	12.53
12	.13	-.12	16.61	21.58	11.64	14.82	-1.79	-.69	18.65	2.49	-3.28	18.20
13	.14	-.02	27.00	20.37	17.88	18.61	7.36	-4.05	17.65	4.56	-.81	19.49
14	.19	-.06	22.99	15.65	18.47	20.83	9.70	-6.03	15.81	2.26	.11	18.24
15	.38	.25	31.47	14.14	26.47	15.88	6.74	1.92	13.33	2.50	5.73	15.01
16	.31	-.11	28.61	15.07	12.03	15.00	16.51	-9.74	10.91	8.29	-1.50	15.03
17	.16	-.16	15.92	11.53	11.90	16.43	8.92	-6.73	11.75	2.01	-1.04	13.98
18	.33	.16	38.35	19.32	30.44	22.04	10.63	-1.11	18.02	3.95	4.88	20.68
19	.33	-.10	105.65	53.22	45.52	55.63	62.54	-36.33	38.79	30.07	-4.45	54.43

Note: v(g) = var(g), etc.; N_M, N_D = sample sizes (Number of pairs).

TABLE 1. A: Jenkins' personality questionnaire.

FEMALES (JF, N_M = 86, N_D = 74)

	Correlations		Mean squares				Parameter estimates					
								AG3			E3	
	r_M	r_D	MSBM	MSWM	MSBD	MSWD	v(a)	v(e)	v(z)	$v(e_M)$	$v(e_D)$	v(z)
1	.03	.20	19.02	20.20	22.65	15.10	-8.73	8.14	19.83	-1.81	2.50	17.65
2	.23	.02	27.64	17.30	19.19	18.44	9.59	-4.42	15.47	4.22	4.22	17.87
3	.30	.37	40.76	21.95	31.97	14.70	1.54	7.87	17.94	4.39	6.82	18.33
4	.16	.35	46.90	33.96	29.89	14.39	-2.56	9.03	24.81	8.51	2.86	24.17
5	.28	.12	45.37	25.52	23.09	18.14	14.90	-4.97	18.10	11.14	.63	21.83
6	.29	-.06	29.89	16.45	17.14	19.33	15.63	-8.91	13.98	6.38	-.38	17.89
7	.18	.03	37.40	25.99	33.66	31.70	9.45	-3.75	26.48	1.87	2.41	28.85
8	.53	-.02	26.14	8.03	16.52	17.19	18.78	-9.73	7.91	4.81	1.96	12.61
9	.47	.37	23.33	8.41	27.25	12.53	.20	7.26	10.42	-1.96	8.39	10.47
10	.25	.15	21.87	13.12	14.65	10.83	4.93	-.56	10.74	3.61	1.34	11.98
11	.40	.04	32.11	13.76	24.36	22.49	16.48	-7.30	14.01	3.88	3.12	18.13
12	.23	.25	22.25	13.93	26.72	16.03	-2.37	6.53	15.57	-2.23	5.87	14.98
13	.42	.28	33.71	13.77	26.15	14.71	8.50	1.47	12.12	3.78	5.95	14.24
14	.12	.17	23.79	18.69	25.47	18.07	-2.30	4.85	18.95	-.84	3.54	18.38
15	.25	-.19	25.83	15.50	13.75	20.20	16.78	-11.62	13.66	6.04	-2.05	17.85
16	.17	.22	26.05	18.48	24.38	15.59	-1.22	5.00	17.34	.84	3.67	7.03
17	.29	-.06	20.93	11.52	12.29	13.86	10.98	-6.27	9.94	4.32	-.20	12.69
18	.49	.22	31.61	10.82	16.56	10.59	14.82	-4.43	7.00	7.53	2.93	10.70
19	.17	.17	76.26	54.10	52.66	37.36	6.86	4.22	44.01	11.80	3.47	45.73

Note: v(g) = var(g), etc.; N_M, N_D = sample sizes (Number of pairs).

TABLE 1. B: Cattell's High School Personality Questionnaire.

	Correlations		Mean squares				Parameter estimates					
								AG3			E3	
	r_M	r_D	MSBM	MSWM	MSBD	MSWD	v(a)	v(e)	v(z)	$v(e_M)$	$v(e_D)$	v(z)
WHITES (HW, N_M = 34, N_D = 29)												
1	.44	-.11	19.26	7.49	13.73	17.12	15.16	-9.28	8.51	2.77	.71	12.31
2	.35	.55	3.88	1.87	4.86	1.41	-1.44	2.45	2.00	-.49	1.61	1.64
3	.21	.21	14.50	9.47	14.55	9.50	-.02	2.54	9.49	-.03	2.53	9.49
4	.21	.34	13.34	8.71	20.40	10.05	-5.72	8.03	10.81	-3.53	5.51	9.38
5	.39	.04	10.46	4.59	10.44	9.64	5.07	-2.14	5.85	.01	1.66	7.12
6	.50	-.16	15.63	5.21	9.52	13.14	14.04	-8.83	5.66	3.05	.17	9.18
7	.44	-.01	12.83	4.99	11.11	11.33	8.06	-4.14	6.14	.86	1.47	8.16
8	.13	.04	15.35	11.82	14.73	13.60	2.40	-.64	12.11	.31	1.01	12.71
9	.74	.67	48.65	7.27	46.96	9.28	3.70	16.99	7.36	.85	19.34	8.27
10	.02	.08	7.03	6.75	10.96	9.34	-1.34	1.48	8.38	-1.96	1.46	8.05
11	-.09	.16	10.34	12.38	15.04	10.89	-6.19	5.17	13.18	-2.35	1.70	11.64
12	.32	.24	15.30	7.88	18.73	11.48	.17	3.54	9.64	-1.71	4.53	9.68
13	.29	-.11	12.72	7.00	11.36	14.17	8.53	-5.67	8.45	.68	.39	10.59
14	.27	.06	16.09	9.25	14.10	12.50	5.24	-1.82	9.57	.99	1.61	10.88
BLACKS (HB, N_M = 48, N_D = 32)												
1	.36	.09	13.69	6.44	10.82	9.03	5.46	-1.83	6.37	1.43	1.54	7.74
2	.48	.18	5.46	1.92	5.89	4.09	1.74	.03	2.57	-.21	1.44	3.01
3	.41	.06	17.49	7.32	11.99	10.63	8.71	-3.73	6.77	2.75	1.51	8.98

4	.19	.04	14.02	9.54	9.21	8.50	3.77	-1.53	8.08	2.41	.10	9.02	
5	.15	.18	13.77	10.18	12.99	9.03	-.37	2.16	9.70	.39	1.69	9.60	
6	.41	.26	13.72	5.74	14.05	8.25	2.18	1.81	6.45	-.16	3.53	6.99	
7	.13	-.11	10.47	8.06	6.70	8.36	4.07	-2.86	7.19	1.89	-.76	8.21	
8	.21	-.26	13.29	8.68	7.58	12.91	9.94	-7.63	8.31	2.86	-1.61	10.80	
9	.45	.41	21.22	8.05	16.87	7.06	3.36	3.23	6.72	2.17	4.66	7.56	
10	.21	.11	10.28	6.71	10.15	8.14	1.56	.22	7.03	.07	1.36	7.43	
11	.27	.15	14.63	8.41	9.88	7.30	3.64	-.53	6.94	2.38	1.01	7.86	
12	.33	.11	11.69	5.89	12.88	10.33	3.25	-.35	7.30	-.60	2.39	8.11	
13	.14	.06	11.67	8.80	9.06	8.03	1.84	-.40	7.95	1.30	.32	8.41	
14	.25	.00	13.52	8.11	10.78	10.78	5.41	-2.71	8.09	1.37	.67	9.44	
TOTAL (HT, N_M = 82, N_D = 61)													
1	.43	-.01	17.24	6.87	12.62	12.88	10.63	-5.45	7.22	2.31	1.37	9.88	
2	.54	.44	6.36	1.90	7.25	2.82	.03	2.20	2.35	-.44	2.44	2.36	
3	.33	.17	16.30	8.21	14.22	10.09	3.96	.09	8.16	1.04	2.54	9.15	
4	.20	.23	13.80	9.20	14.76	9.24	-.92	3.22	9.45	-.48	2.77	9.22	
5	.23	.12	12.56	7.86	11.86	9.32	2.16	.19	8.05	.35	1.63	8.59	
6	.51	.12	17.01	5.52	13.45	10.57	8.61	-2.87	5.89	1.78	2.70	8.05	
7	.26	-.05	11.56	6.79	8.84	9.77	5.70	-3.32	6.86	1.36	.28	8.28	
8	.17	-.08	14.07	9.98	11.28	13.24	6.05	-4.01	10.10	1.39	-.16	11.61	
9	.63	.59	34.05	7.73	31.45	8.11	2.98	10.18	7.18	1.30	11.77	7.92	
10	.15	.11	9.11	6.73	10.86	8.71	.23	.96	7.66	-.88	1.57	7.72	
11	.12	.16	12.79	10.05	12.50	9.05	-.71	2.08	9.73	.14	1.47	9.55	
12	.33	.22	13.32	6.71	17.02	10.88	.47	2.84	8.68	-1.85	4.11	8.80	
13	.20	-.04	12.08	8.05	10.11	10.95	4.87	-2.86	8.28	.99	.30	9.50	
14	.26	.03	14.63	8.59	12.32	11.60	5.32	-2.30	8.77	1.16	1.11	10.10	

TABLE 2.

Heritability estimates HR for Osborne's personality data.

Variable	JM	JF	HW	HB	HT
1	2.54	15.34	2.50	1.50	2.04
2	1.30	1.82	-1.14	1.26	.38
3	2.58	-.46	.00	1.70	.96
4	.60	-2.38	-1.24	1.58	-.30
5	2.12	1.14	1.80	-.40	.96
6	1.24	2.42	2.64	.74	1.52
7	2.44	1.66	2.04	3.70	2.38
8	1.00	2.08	1.38	4.48	2.94
9	2.22	.42	.18	.18	.12
10	.40	.80	-6.00	.96	.54
11	1.82	1.80	5.56	.88	-.66
12	.16	-.18	.50	1.34	.66
13	2.28	.66	2.76	1.14	2.40
14	2.64	-.84	1.56	2.00	1.76
15	.68	3.52			
16	2.70	-.58			
17	4.00	2.42			
18	1.04	1.10			
19	2.60	.00			

Distribution of heritability ratios HR

Study	n	< 0	0-1	> 1
JM	19	.000	.263	.789
JF	19	.263	.211	.526
HW	14	.214	.214	.571
HB	14	.071	.286	.643
HT	14	.143	.429	.429
ALL	80	.138	.263	.589

Note: n = number of variables. Since the heritability ratios HR were computed on the basis of the reconstructed MSB, some of them differ slightly from those reported by Osborne (1980, p. 127, p. 135f) due to rounding error. The distributions are the same.

As noted above, a case can be made only for HR, since Holzinger's h^2 is not a valid heritability estimate under any conditions. However, the case for HR remains cogent only as long as the model from which it is derived, the model AG3, describes the data. The high proportion of inadmissible HR already suggests that this genetic model may not describe Osborne's personality data very well.

Qualitative comparison of AG3 with E3: Closer scrutiny of the violations of the qualitative restrictions implied by AG3 confirms this first impression. These violations are tabulated in Table 3.

TABLE 3. Proportions of violations of ordinal predictions.

	$r_M, r_D < 0$	MSWD< MSWM	$r_D < r_M/2$	MSBM+MSWM <MSBD+MSWD
		Observed		
JM	.32	.37	.79	.21
JF	.13	.47	.53	.26
HW	.29	.14	.71	1.00
HB	.07	.36	.61	.34
HT	.14	.07	.43	.79
ALL	.18	.30	.59	.49
		Predicted		
AG3	0	0	0	.5
E3	0	.5	-	0

(a) Column 1 in Table 3 gives the proportion of violations of the weak conditions $\rho_M \geq 0$, $\rho_D \geq 0$, which apply to both AG3 and E3. Overall, they are violated in 18% of all 160 cases. Most of the violations occur for r_D in the JM group, suggesting that, for most of the variables, ρ_D is zero in this group.

(b) Column 2 gives the relative frequencies of negative values for the difference MSWD-MSWM. Under AG3, this difference should be non-negative, since it estimates var(a)/2. Under E3 it should be zero, because both MSWM and MSWD estimate the same error variance, var(z). Overall, this difference is negative in 30% of all 80 cases, a percentage halfway between 0, as predicted by AG3, and 50, as predicted by E3.

(c) The model AG3 predicts that MSB+MSW should be the same for both twin types, while model E3 predicts that it should be larger for the MZs. The average proportion of MSBM+MSWM < MSBD+MSWD (column 4), 49%, is misleading, because of systematic violations of the equality prediction within subgroups. For both sets of the Jenkins data, MSBM+MSWM > MSBD+MSWD in 3/4 of all cases, favoring E3. In the HB group, this percentage is 64%. In the HW group, it is one, contradicting both AG3 and E3. If one takes the average magnitude of the contrasts (MSBD+MSWD)-(MSBM+MSWM) into account, this comparison favors E3: For JM and JF this average is -11.6 and -9.1. For the three HSPQ sets, it is 3.6, -1.2, and 1.2. Allowing for the fact that the mean squares for the Jenkins data are on average twice as large as those for the HSPQ, one finds that the equality predictions of AG3 are more severely violated by the Jenkins data in the direction predicted by E3 than the ordinal predictions of E3 by the HSPQ data.

(d) The strong constraint $r_D > r_M/2$ that AG3 imposes on the data is violated in 59% of the cases overall (column 3 of Table 3). Concretely, this means that MZs are more similar to each other than AG3 allows. Since E3 is non-committal on the relative magnitudes of r_M and r_D, this outcome also favors E3.

All in all, these preliminary qualitative checks suggest that the environmental model E3 will fit the Jenkins data much better and the HSPQ data at least moderately better than does the genetic model AG3.

Quantitative comparison of AG3 with E3: To check this conjecture, AG3 and E3 were fitted to all 5 data sets by least squares with the design matrices given in Appendices A3 and A4. The resulting least squares estimates (LSEs) for var(a), var(e), var(z) under AG3 are presented in columns 7-9 of Table 1; those of var(e_M), var(e_D), var(z) under E3 in columns 10-12. Inspection of the LSEs reveals that many more variance estimates are negative under AG3 than under E3. The proportions of inadmissible variance estimates are summarized in Table 4.

In particular, the majority (60%) of the LSEs of the environmental variance component, var(e), are negative. Inspection of the actual LSEs in Table 1 shows that they are often substantially negative. The average magnitude of the negative estimates of var(e) is 5.39. This echoes similar findings reported in (Schönemann, 1990) for Shields' MZA/MZT data which included two cognitive tests. In the present MZ/DZ study,

25% of all variance estimates are negative under AG3, compared to only 14% under the environmental model E3. Similarly, the average magnitude of all negative variance estimates is 3.68 under AG3, while E3 reduces it to 1.17 (Table 4). While small negative LSEs are to be expected when the underlying parameter is close to zero, the consistently negative estimates of the magnitude encountered with these data strongly suggest that the model AG3 underlying HR is inappropriate for these data.

TABLE 4. Proportion of inadmissible (negative) variance estimates.

		Genetic Model AG3			Environmental Model E3		
Study	n	var(a)	var(e)	var(z)	var(e_M)	var(e_D)	var(z)
JM	19	.05 (1.79)	.79 (8.75)	.00	.11 (12.51)	.47 (1.60)	.00
JF	19	.26 (3.44)	.53 (6.20)	.00	.21 (1.71)	.16 (.88)	.00
HW	14	.36 (2.94)	.50 (4.64)	.00	.43 (1.68)	.00	.00
HB	14	.07 (.37)	.64 (2.40)	.00	.21 (.33)	.14 (1.18)	.00
HT	14	.14 (.81)	.43 (3.46)	.00	.29 (.91)	.07 (.16)	.00
ALL	80	.20 (1.96)	.59 (5.39)	.00	.24 (1.51)	.19 (.82)	.00
Average magnitude over studies and parameters		(3.68)			(1.17)		
Fit ratio AG3/E3			3.15				

Note: average magnitude of negative variance estimates in parentheses.

Overall, in terms of proportion of negative variance estimates, the environmental model fits better by a factor of 1.8. In terms of the average magnitude of negative variance estimates, E3 fits better by a factor of 3.2 (Table 4).

Quantitative fit: Since a variance cannot be negative by definition, the only inference a negative variance estimate permits is that the underlying population parameter is zero. Hence, before the relative fit of both models can be evaluated, all negative variance estimates must first be replaced by zero.

TABLE 5. Comparison of fit of the genetic model AG3 versus the environmental model E3.

Variable	Study									
	JM		JF		HW		HB		HT	
	e_G	e_E	e_G	e_E	e_G	e_E	e_G	e_E	e_G	e_E
1	.64	.01	.33	.02	.76	.05	.06	.01	.36	.03
2	.30	.01	.10	.00	.31	.02	.02	.03	.01	.01
3	.03	.00	.02	.01	.00	.00	.18	.01	.00	.00
4	.03	.04	.08	.04	.30	.07	.06	.00	.01	.00
5	.38	.00	.12	.01	.13	.04	.00	.00	.00	.00
6	.12	.00	.36	.00	1.17	.06	.00	.01	.11	.02
7	.48	.00	.03	.00	.32	.05	.24	.02	.25	.01
8	.00	.00	.57	.03	.00	.00	.98	.06	.21	.01
9	.23	.00	.01	.02	.00	.00	.01	.00	.00	.00
10	.01	.00	.03	.00	.07	.06	.00	.00	.01	.02
11	.53	.00	.18	.02	.42	.04	.02	.00	.01	.00
12	.07	.10	.03	.01	.02	.02	.02	.03	.02	.04
13	.09	.00	.00	.00	.49	.05	.01	.00	.15	.01
14	.19	.01	.02	.00	.04	.01	.12	.01	.07	.01
15	.00	.00	.73	.03						
16	.59	.01	.01	.00						
17	.45	.03	.36	.00						
18	.01	.00	.14	.00						
19	.59	.01	.03	.01						
ALL	.25	.01	.17	.01	.29	.03	.12	.01	.09	.01
Relative fit ratios: e_G/e_E	25.0		17.0		9.7		12.0		9.0	

Overall fit ratio across all tests and groups: $e_G/e_E = 13.9$

Notes: $e_G = e_G'e_G/y'y = 1 - \eta^2$ under AG3
$e_E = e_E'e_E/y'y = 1 - \eta^2$ under E3

i.e., e_G is the residual vector after the least squares fit of AG3 to the mean square criterion vector y, and e_E is the LS residual vector after fitting E3 to the mean squares in y, after all negative variance estimates were replaced by zero.

Once this is done, one obtains the $1-\eta^2$ values in Table 5 ($1-\eta^2 := e'e/y'y$, where now y denotes the criterion vector of the mean squares, and e the least square residual vector). These values are abbreviated "e_G" under AG3, and "e_E" under E3. In terms of the fit ratios e_G/e_E, the environmental model E3 fits the JM data better than the genetic model by a factor of 25, and the JM data better by a factor of 17. For HW, HB, and HT, the fit ratios favoring the purely environmental model E3 are 9.7, 12, and 9. The average fit ratio e_G/e_E over all 5 studies is 13.9.

5. DOMINANCE MAY INFLATE HR

As was shown in section 2, Nichols' HR is a valid deduction from the strictly additive genetic model AG3. However, we found empirically (a) that this model does not fit Osborne's data, and (b) that HR tends to be biased upward, i.e., an excessive proportion of HRs exceed 1. In this section, an attempt will be made to reconcile these two findings by embedding the strictly additive model AG3 into a more comprehensive genetic model which provides for dominance.

This model is a variant of a model Fisher developed in (1918) to separate additive gene action from dominance and other non-additive genetic factors (see, e.g., Harris, 1965; Falconer, 1960; McGuire & Hirsch, 1977). If one assumes that all interactions and the gene/environment correlation are zero and adds measurement error, then one arrives at a variant of Fisher's model, DG4, which contains 4 parameters:

(a) the variance, var(a), of the additive values a of gene action,
(b) the variance, var(d), of the dominance deviations,
(c) the environmental variance, var(e), and
(d) the variance of the measurement error, var(z).

Fisher partitioned the total genetic variance, var(g), into two components, the additive part, var(a), and the dominance part, var(d). The notion of dominance is most easily understood in the special case of one gene locus with two alleles, a and A, giving rise to three genotypes, aa, aA, and AA. Disregarding all non-genetic factors, if the phenotypic expression y of the heterozygotes aA is exactly halfway between those of the two homozygotes, aa and AA (so that $|\, y(aa)-y(aA)\,| = |\, y(AA)-y(aA)\,|$ if $|\, aa-aA\,| = |\, AA-aA\,|$ on the abscissa), then gene action is called "additive". For example, if a is the gene for white and A the gene for red, then y(aA), the color of a cross between aa and AA flowers may be pink. If y(aA) is reddish pink, then A is "partially dominant" over white, and if y(aA) = y(AA), i.e., pure red again, it is said to be "completely dominant."

To measure the degree of dominance, Fisher fitted a regression line to the three phenotypic expressions, y(aa), y(aA), and y(AA) (see, e.g., Harris, 1965, p. 84, and MacGuire & Hirsch, 1977, for a numerical illustration). The three predicted values a_k on this line are the "additive genetic values" and the three deviations from this line are the "dominance deviations", d_k. As a consequence of the least squares definition, the additive values a are uncorrelated with the dominance deviations d, so that $g := a + d$ implies $var(g) = var(a) + var(d)$.

This reasoning can be extended to $N > 1$ loci with more than two alleles. However, in this case, two loci may interact.

If one is willing to ignore this and other complications, then the partition of the total genetic variance var(g) can also be computed for relatives with partial genetic overlap due to various degrees of common ancestry, such as fraternal twins, half-sibs, nephews, etc. The genetic covariance $cov(g_1,g_2)$ for two relatives can then be written as a linear combination of var(a), var(d):

$$cov(g_1,g_2) = w_a var(a) + w_d var(d),$$

where the "co-ancestral coefficients" (Malecot, 1969; Harris, 1965, p. 86) w_a, w_d reflect the degree of genetic overlap. As long as all other variables entering the observed phenotypic values y_1, y_2 are uncorrelated with a and d (and hence g), this covariance equals the phenotypic covariance, $cov(y_1,y_2)$.

For MZs, these coefficients are $w_a = w_d = 1$, because they share all genes. Hence, for MZs, the model DG4 postulates

$$\text{MZs:}\quad y_1 - \mu = a + d + e + z_1, \qquad y_2 - \mu = a + d + e + z_2$$

with $E(a,d,e,z_1,z_2) = 0$, and

$$Var(a,d,e,z_1,z_2) = diag[var(a),var(d),var(e),var(z_1),var(z_2)],$$

(left side of Fig. 1C). By A1, the expected mean squares for the MZs are, therefore,

$$E(MSBM) = 2var(a)+2var(d)+2var(e)+var(z),\ E(MSWM) = var(z),$$

and the intraclass correlation is

$$\rho_M = [var(a)+var(d)+var(e)]/[var(a)+var(d)+var(e)+var(z)].$$

For fraternal twins, the co-ancestral weights are $w_a = 1/2$, $w_d = 1/4$ (see Harris, loc. cit.), so that

$$cov(g_1,g_2) = cov(a_1+d_1,a_2+d_2) = var(a)/2 + var(d)/4.$$

Hence, the model for the DZs is:

$$\text{DZs:}\quad y_1 - \mu = a_1 + d_1 + e + z_1, \qquad y_2 - \mu = a_2 + d_2 + e + z_2,$$

with $cov(a_1,a_2) = var(a)/2$, $cov(d_1,d_2) = var(d)/4$, while all other variables remain uncorrelated (right side of Fig. 1C).

Therefore,

$$E(MSBD) = 1.5var(a) + 1.25var(d) + 2var(e) + var(z)$$
$$E(MSWD) = .5var(a) + .75var(d) + var(z),$$
$$\rho_D = [.5var(a)+.25var(d)+var(e)]/[var(a)+var(d)+var(e)+var(z)].$$

Since $g = a+d$ is uncorrelated with e and z, one also finds

$$cov(g_1,g_2) = cov(y_1,y_2) = E(MSB\text{-}MSW)/2.$$

On making this substitution, one arrives at the expected within mean squares and covariances given by Falconer (1960, p. 184).

One again finds that Holzinger's h^2 (and, hence, Vandenberg's F, which is simply a monotone function of it) estimates nothing of value, because it does not depend on var(e).

On inspecting the design matrix for this model,

$$\begin{pmatrix} 2 & 2 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 1.5 & 1.25 & 2 & 1 \\ .5 & .75 & 0 & 1 \end{pmatrix}$$

one finds that it has rank 3, since the sum of the first two rows equals the sum of the last two rows. Hence, the parameters var(a), var(d), var(e), and var(z) are not estimable from MZ/DZ data alone. However, the linear combinations

var(z), var(a) + 1.5var(d), and var(a) + var(d) + var(e),

are estimable, because their coefficient vectors are in the row space of the design matrix. To estimate all four parameters of this model would require additional data from other relatives, e.g., half-brothers or half-sisters, or adopted children raised in the same environment, so that the augmented design matrix acquires full column rank.

For our present purposes, it suffices to note that this model contains the strictly additive model AG3 as a special case for var(d) = 0. This can be seen algebraically by deleting the second column of the design matrix of DG4, and graphically by deleting the broken arrows in the two diagrams for DG4 (Fig. 1C), which leads back to AG3 (Fig. 1A). Under DG4, one finds for Nichols' HR,

$$HR := 2(\rho_M - \rho_D)/\rho_M = var(a)+1.5var(d)]/[var(a)+var(d)+var(e)]$$

which is larger than

$$var(g)/[var(g)+var(e)] =: h_B,$$

the "broad heritability" (where $g := a+d$) if var(d) > 0.

In other words, while Nichols' HR is the correct expression of narrow (and hence broad) heritability when there is no dominance (i.e., under AG3), it will *overestimate* broad heritability if there is partial dominance. This might explain the consistently upward bias of HR.

6. DISCUSSION

We are thus left with the overall conclusion that the strictly additive genetic model AG3 needed to justify Nichols' heritability estimate HR is incompatible with Osborne's data, so that the HRs are empirically meaningless. The consistently better fit of the purely environmental model E3 could mean that most personality tests contain no genetic components at all, confirming Jones' (1946) hypothesis that the greater similarity of the MZs relative to the DZs on personality tests is due to additional environmental variance induced by the greater physical resemblance of MZs. Because they look more alike, they are treated more alike than the DZs. While a descriptive variance component model can never prove such a causal relationship, Osborne's personality data are at least consistent with it.

Although these specific conclusions are limited to Osborne's personality data, the basic principles we used to arrive at them are relevant more generally: Since all descriptive models are to a large extent arbitrary, it is critical that the underlying assumptions are made explicit so that they can be tested empirically. Failure to do so engenders the risk of interpreting parameter estimates (e.g., HRs) which are empirically meaningless, and hence substantively misleading, because the data do not fit the model from which they were derived: "Unfortunately many...working on IQ are so involved in the model-game that they pay little attention to the data they put in them" (Roubertoux & Capron, 1990, p. 564; see also Kempthorne, 1990, p. 139).

In the present case, we found that most of the assumptions of the variance component model needed to justify HR as a valid heritability estimate were violated by Osborne's data. In particular, we found that the prediction $r_D \geq r_M/2$ was often violated, i.e., concretely, that the similarity of the MZs exceeds that of the DZs to a larger degree than the genetic model allows. Consequently, most of the estimates of var(e) were negative and many HRs exceeded 1. If we had ignored these systematic violations, we might have concluded that narrow heritability is 1 for most of the personality tests. However, once we recognized the violations and found that a purely environmental model fit these data much better, we were led to the opposite conclusion, namely that broad heritability is zero for these data.

Of course, these general principles also apply to heritability ratios of IQs. So far as we know, the popular belief that 80% of the IQ variance is genetic rests on heritability estimates unsupported by any evidence

that the underlying models actually fit the data. Unless these assumptions are checked, and the fit of genetic models is fairly compared with the fit of competing environmental models, no one can know whether the heritability figures for IQ are any more valid than those for personality tests. (see also Schönemann, 1989, 1990; Schönemann & Schönemann, 1991).

ACKNOWLEDGEMENTS

This work was completed while the senior author held a Visiting Research Professorship at the National Taiwan University, Taipei, ROC. The support of the National Science Council is gratefully acknowledged.

REFERENCES

Cattell, R. B. (1958). *Handbook for the IPAT High School Personality Questionnaire.* Champaign, IL: IPAT.

Cattell, R. B. (1969). *Junior High School Personality Questionnaire.* Champaign, IL: IPAT.

Falconer, D. S. (1960). *Introduction to quantitative genetics.* Edinburgh and London: Oliver and Boyd.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52,* 399-433.

Harris, D. L. (1965). Biometrical genetics in man. In S. Vandenberg (Ed.), *Methods and goals in human behavior genetics.* New York: Academic Press.

Hearnshaw, L. S. (1981). *Cyril Burt, Psychologist.* New York: Vintage.

Hogben, L. (1950). *Chance and choice by cardpack and chessboard* (Vol. I). London: Parrish.

Jenkins, T. N. (1959). *How well do you know your self?* New York: Executive Analysis Corporation.

Jensen, A. R. (1967). Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proceedings of the National Academy of Sciences, 58,* 149-159.

Jensen, A. R. (1970). Statement of Dr. Arthur R. Jensen before the General Subcomittee on Education, House Education and Labor Committee. Washington, Congressional Record, I-237.

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, H. E. (1946). Environmental influences on mental development. In L. Charmichael (Ed.), *Manual of child psychology.* New York: Wiley. [Fourth edition 1949.]

Jones, H. E., & Wilson, P. T. (1932-1933). Reputation differences in like-sex twins. *Journal of Experimental Education, 1,* 86-91.

Kamin, L. J. (1974). *The science and politics of IQ.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Kempthorne, O. (1990). How does one apply statistical analysis to our understanding of the development of human relation ships? *Behavioral and Brain Sciences, 13,* 138-39.

King, J. C. (1981). *The biology of race.* Berkeley, CA: University of California Press.

Lehtovaara, A. (1938). *Psychologische Zwillingsuntersuchungen.* Helsinki: Academiae Scientiarium Fennicae.

Loehlin, J. C., & Nichols, R. C. (1976). *Heredity, environment, & personality.* Austin & London: University of Texas Press.

Malecot, G. (1969). *The mathematics of heredity.* San Francisco: Freeman.

McGuire, T. R., & Hirsch, J. (1977). General intelligence (g) and heritability (H^2, h^2). In I. Uzgiris & F. Weizman (Eds.), *The structure of experience.* New York: Plenum.

Newman, H., Freeman, F., & Holzinger, K. (1937). Twins. *A study of heredity and environment.* Chicago: University of Chicago Press.

Nichols, R. C. (1965). The National Merit twin study. In G. S. Vandenberg (Ed.), *Methods and goals in human behavior genetics.* New York: Academic Press.

Osborne, T. (1980). *Twins: Black and white.* Athens, GA: Foundation for Human Understanding.

Roubertoux, P. L., & Capron, C. (1990). Are intelligence differences heridi-tarily transmitted? *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology, 10,* 555-594.

Rowe, D. C. (1981). Environmental and genetic influences of dimensions of perceived parenting: A twin study. *Developmental Psychology, 17,* 203-208.

Schönemann, P. H. (1989). New questions about old heritability estimates. *Bulletin of the Psychonomic Society, 27,* 175-178.

Schönemann, P. H. (1990). Environmental versus genetic variance component models for identical twins: A critique of Jinks and Fulker's reanalysis of the Shields data. *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology, 10,* 451-473.

Schönemann, P. H., & Schönemann, R. D. (1991). We wondered where the errors went. *Behavioral and Brain Sciences, 14,* 404-406.

Shields, J. (1962). *Monozygotic twins brought up apart and brought up together.* London: Oxford University Press.

Stocks, P. (1930-1931). A biometric investigation of twins and their brothers and sisters. *Annals of Eugenics, 4,* 49-108.

Wilson, P. T. (1934). A study of twins with special reference to heredity as a factor determining differences in environment. *Human Biology, 6,* 324-354.
Wright, S. (1978). *Evolution and the genetics of populations. Volume 4: Variability within and among populations.* Chicago: The University of Chicago Press.

APPENDIX 1. Twin data, intraclass correlations, and mean squares

If $Y := (y_{ik})$ is an Nx2 table of scores for N twin pairs, and $\bar{y} := \Sigma_k\Sigma_i\, y_{ik}/2N$ is the overall mean, then the conventional sums of squares are given by

$$SSB = \Sigma i\, (y_{i1} + y_{i2} - 2\bar{y})^2/2,\ SSW = \Sigma_i\, (y_{i1} - y_{i2})^2/2,$$

i.e., by half the sums of squares of the row sums and row differences of the deviation score matrix around the grand mean, as was shown in more detail in (Schönemann, 1990).

For the derivations of the expected mean squares, the following well-known result is used repeatedly:

If x is an N-variate random vector with $E(x) = \phi$ and $Var(x) = \sigma^2 I_N$, and $P = P^2 = P'$ is an NxN idempotent matrix of rank r, then the quadratic form $q := x'Px$ has expected value $E(q) = r\sigma^2$. (Since $E(p_{ij}x_ix_j) = 0$ for $i \neq j$, $E(\Sigma_i\, p_{ii}x_i^2) = \sigma^2$ trace(P), and trace(P) = r.)

In the population, an intraclass correlation can be defined in several ways, e.g., in terms of expected mean squares as

$$\rho := E(MSB - MSW)/E(MSB + MSW),$$

which is the definition we use throughout this paper.

APPENDIX 2. Holzinger's derivation of h^2

This section briefly sketches an argument developed more fully in (Schönemann, 1989).

To derive his heritability ratio h^2, Holzinger (Newman, Freeman, & Holzinger, 1937, pp. 94-116) started out with a variance component model for the MZs which postulates 4 latent variables, g (genetic component), e_1, e_2 (environmental components), z_1, and z_2 (measurement errors):

$$\text{MZs:}\quad y_1 - \mu = g + e_1 + z_1, \quad y_2 - \mu = g + e_2 + z_2$$

where $\mu := E(y_k)$, $E(g) = E(e_k) = E(z_k) = 0$ and $Var(g,e_1,e_2,z_1,z_2)$ = diagonal (i.e., all latent variables are uncorrelated) and $var(z_k)$ = var(z), $var(e_k)$ = var(e). However, halfway through his derivations,

Holzinger discarded the measurement error variables by setting var(z) = 0 (p. 113). Therefore, the model he actually worked with predicts

$$E(MSBM) = 2var(g) + var(e), \qquad E(MSWM) = var(e),$$
$$\rho_M = var(g)/[var(g)+var(e)],$$

and the broad heritability h_B is simply given by the intraclass correlation of the monozygotic twins, a fact Holzinger overlooked. Instead, he went on to postulate a variance component model for the DZs to be able to define his heritability ratio as $h^2 := (\rho_M - \rho_D)/(1 - \rho_D)$. After again dropping the measurement error variable z, he described the DZs with

$$\text{DZs:}\quad y_1 - \mu = g_1 + e_1, \quad y_2 - \mu = g_2 + e_2.$$

He again assumed that all variables on the right of the equality signs are uncorrelated within twin types. However, he did not spell out the critical correlation between the two genetic variables g_1, g_2. Suppose it is given by ρ_{12}. Then, by A1:

$$E(MSBD) = (1 + \rho_{12})var(g) + var(e),$$
$$E(MSWD) = (1 - \rho_{12})var(g) + var(e)$$
$$\rho_D = \rho_{12}var(g)/[var(g)+var(e)].$$

This implies that his heritability coefficient can be written

$$h^2 := (\rho_M-\rho_D)/(1-\rho_D) = (1-\rho_{12})var(g)/[(1-\rho_{12})var(g)+var(e)].$$

On setting $h^2 = var(g)/[var(g)+var(e)]$, one obtains $\rho_{12}var(g)var(e) = 0$, so that $\rho_{12} = 0$ if both variances are nonzero. As a result, $\rho_D = 0$. The necessary condition that $\rho_{12} = 0$ violates the basic assumption of genetic theory that dizygotic twins, on average, share half of the genes.

APPENDIX 3. The strictly additive genetic variance component AG3

This model is an adaptation to MZ/DZ data of the model Jinks and Fulker (1970; s.a. Schönemann, 1990), used in their reanalysis of the Shields MZA/MZT data. Following Jinks and Fulker, we will present it here in purely statistical terms. As shown in the text, this model is a variant of Fisher's (1918) model if the gene/environmental correlation, interactions, and dominance are all zero. The model AG3 postulates 3 classes of latent variables for MZs and DZs reared together: strictly additive genetic variables a, environmental variables e, and measurement error z. The structural part for the MZs is

$$\text{MZs:}\quad y_1 - \mu = a + e + z_1, \quad y_2 - \mu = a + e + z_2,$$

with $E(a) = E(e) = E(z_k) = 0$ and $Var(a,e,z_1,z_2)$ = diagonal as stochastic restrictions. Note that now, in contrast to Holzinger's model,

the environmental variable e is constant within families (since both twins are raised together), so that e and z are no longer confounded. By A1, this implies for the MZs

$$E(MSBM) = 2var(a) + 2var(e) + var(z),\ E(MSWM) = var(z),$$
$$\rho_M = [var(a)+var(e)]/[var(a)+var(e)+var(z)].$$

For the DZs, this model postulates

$$DZs:\quad y_1 - \mu = a_1 + e + z_1,\quad y_2 - \mu = a_2 + e + z_2,$$

where the latent variables again have zero means, and

$$Var(a_1,a_2,e_1,e_2,z_1,z_2) = \text{diagonal } \textit{except} \text{ for } cov(a_1,a_2).$$

Since dizygotic twins, on average, share half the genes, the correlation ρ_{12} between a_1 and a_2 is 1/2, by the correlation formula for common elements (Hogben, 1950, p. 360f). Therefore,

$$E(MSBD) = 1.5var(a)+2var(e)+var(z),$$
$$E(MSWD) = .5var(a)+var(z),$$
$$\rho_D = [.5var(a)+var(e)]/[var(a)+var(e)+var(z)],$$

so that $\rho_M-\rho_D = .5var(a)/[var(a)+var(e)+var(z)]$. Hence

$$HR := 2(\rho_M-\rho_D)/\rho_M = var(a)/[var(a)+var(e)],$$

is the narrow heritability ratio under AG3. In contrast, Holzinger's

$$h^2 := (\rho_M-\rho_D)/(1-\rho_D) = .5var(a)/[.5var(a)+var(z)]$$

measures nothing of interest, because it does not contain the environmental variance, var(e).

It is important to note that AG3 implies a number of conditions which should be checked before relying on HR: The relatively weak prediction $0 \le \rho_D \le \rho_M \le 1$ is met by most MZ/DZ data. In contrast, the stronger predictions

$$\text{(a) } E(MSWD-MSWM) = var(a)/2 \ge 0,$$
$$\text{(b) } E(MSBM+MSWM) = E(MSBD+MSWD),$$
$$\text{(c) } E(MSBD-MSWD) - E(MSBM-MSWM)/2 = var(e) \ge 0,$$
$$\text{(d) } \rho_D \ge \rho_M/2,$$

are often violated. When they are, all parameter estimates, including HR, are of course meaningless, because they are derived from a model which does not fit the data.

The linear model for fitting AG3 to the observed mean squares by the method of least squares is

$$\begin{pmatrix} MSBM \\ MSWM \\ MSBD \\ MSWD \end{pmatrix} = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 1.5 & 2 & 1 \\ .5 & 0 & 1 \end{pmatrix} \begin{pmatrix} var(a) \\ var(e) \\ var(z) \end{pmatrix} + \text{error}$$

Since the coefficient matrix has rank 3, the LSEs are unique.

APPENDIX 4. A 3-parameter environmental model E3

To obtain a model which predicts larger intraclass correlations for the MZs than for the DZs, we postulate two environmental variables: an environmental baseline variable e_D which enters the scores of both MZs and DZs, and a second component e_M which contributes only to the scores of the MZs, thus raising their environmental variance. In addition, we again postulate a measurement error variable z. Thus, we postulate for the

$$MZs:\ y_1 - \mu = e_M + e_D + z_1,\quad y_2 - \mu = e_M + e_D + z_2$$
$$E(e_M,e_D,z_1,z_2) = \phi',$$
$$Var(e_M,e_D,z_1,z_2) = \text{diagonal with } var(z_1) = var(z_2),$$
$$DZs:\ y_1 - \mu = e_D + z_1,\quad y_2 - \mu = e_D + z_2,$$

with the same stochastic restrictions on the remaining variables as for the MZs. By A1, this model implies

$$E(MSBM) = 2var(e_M) + 2var(e_D) + var(z),\ E(MSWM) = var(z),$$
$$E(MSBD) = 2var(e_D) + var(z),\ E(MSWD) = var(z),$$

and, hence,

$$E(MSBM+MSWM) \ge E(MSBD+MSWD)$$
$$E(MSBM-MSBD) = 2var(e_M) \ge 0,$$
$$E(MSWM) = var(z) = E(MSWD),$$
$$1 \ge \rho_M = [var(e_M)+var(e_D)]/[var(e_M)+var(e_D)+var(z)]$$
$$\ge var(e_D)/[var(e_D)+var(z)] = \rho_D \ge 0.$$

The linear model is

$$\begin{pmatrix} MSBM \\ MSWM \\ MSBD \\ MSWD \end{pmatrix} = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} var(e_M) \\ var(e_D) \\ var(z) \end{pmatrix} + \text{error}$$

The LSEs are again unique.

Submitted July 2, 1992
Accepted March 18, 1993