CPC, 1997, 16 (6), 665-694

Famous artefacts: Spearman's hypothesis

Peter H. Schönemann

National Taiwan University

Abstract. In a number of publications, Jensen has recalled Spearman's (1927, p. 379) observation that the loadings of the first principal component (PC1) of various 'intelligence tests' tend to correlate positively with the corresponding Black/White mean differences ('Spearman's Hypothesis'). Jensen believes this sheds light on the true nature of g, Level II Ability, test bias, and Black/White differences. His claims have been warmly welcomed in some quarters (most recently by Herrnstein and Murray, 1994) as conclusive confirmation of the Black inferiority myth. Here it is shown by way of empirical, numerical, geometric, and algebraic demonstrations that the positive correlations predicted by Spearman's hypothesis are psychometric artefacts which also arise (a) with measures which have nothing to do with 'general ability', for example, the number of toys and books a child has; and, more generally, (b) with any set of moderately correlated random data, once the sample is split into high and low groups. Specifically, this interpretation predicts that if sample sizes differ substantially, then the correlation will be larger for the PC1 of the larger group. This prediction is borne out both in simulated and in 'real' data sets, including Jensen's.

Key words: Spearman hypothesis, Spearman g, artefacts, Black/White IQ differences, group differences on mental tests, racism.

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 his book, *Bias in mental testing*, Jensen (1980) devoted consider-
le space to « Spearman's interesting hypothesis that the magnitude of
hite-Black mean differences on various mental traits are directly re-
ted to the test's g loadings (Spearman, 1927, p. 379) » (Jensen, 1980,
. 535. See also pp. 541, 544-545, 548-554, 585-586, 732 ff.)

In a target article for *The Behavioral and Brain Sciences* (Jensen,
985), he expanded on this theme, pointing to the results of his reanal-
sis of 11 data sets as evidence « that the varying magnitude of the mean
ifference between Black and White populations on a variety of mental
ests is directly related to the size of the test's loading on g, the general
actor common to all complex tests of mental ability ... in accord with
pearman's hypothesis ... » (ibid., Abstract). Jensen believes « An im-
ortant practical implication of Spearman's hypothesis » is that « Black
eople, statistically, will have a greater handicap in those educational,
ccupational, and military assignments that are most likely correlated
vith measures of general intelligence ... The practical implications of g
nd Spearman's hypothesis for employment, productivity, and the na-
ion's economic welfare have been discussed in more detail elsewhere. »

Numerous commentators on Jensen's BBS target article, and subse-
quently, Gordon in Modgil's book on Jensen (Modgil, 1987, pp. 113-
114, 120-139, 141-1422, 144, 172, 205, 207, 357-358, 361, 368)
seconded Jensen's bold conjectures:

Brandt (1985, p. 222): « Jensen's [BBS] article is both scholarly and
powerful: with all the skill and assiduity of the world's most impressive
psychometrician he mounts an argument that should subdue objection
and compel assent. »

Carlson (1985, p. 224): « In his assessment of Spearman's hypothesis
Jensen provides a valuable and scholarly review of research and theory
concerning g. »

Cattell (1985, p. 227): « Jensen sets out, with impeccable scientific
method, to supply the first possible alternative corroboration one would
want to see, to his finding of significant intelligence differences between
Blacks and Whites. »

Jones (1985, p. 233): « Arthur Jensen's reanalysis of data from 11
studies provides convincing evidence that the observed differences be-
tween average scores of Black and White samples in the United States in
a variety of mental tests are directly related to average differences in g.»

Kline (1985, p. 234): «It is difficult to impugn the logic of the procedure: Positive correlations support the claim; anything else refutes it. In fact, the hypothesis was entirely and strongly supported. »

Nettlebeck (1985, p. 235): « Jensen's extensive analysis confirms Spearman's suggestion that significant mean differences in IQ between Blacks and Whites in the US reflect differences in g. »

Nichols (1985, p. 236): « The empirical support for this hypothesis reported by Jensen seems more than adequate. In fact, the evidence is so strong and pervasive that the impressive technical sophistication of the analysis hardly seems necessary. »

Gordon (1987, p. 122): « The Spearman hypothesis ... was potentially a very important hypothesis indeed, perhaps even the single most important hypothesis to emerge from Jensen's research on bias in view of its scope and hence its capacity for tying together many critical but poorly understood phenomena ». In deference to the historic significance of Jensen's contribution, Gordon also records that « Jensen first attempted to publish analyses concerning the Spearman hypothesis in 1979 in the *American Psychologist*. After an unusually long review process of eleven months, Jensen's paper was rejected, with no encouragement to revise or resubmit, despite recommendations to that effect from two referees and a favorable review from a third » (ibid).

In his concluding remarks, Jensen (1987, p. 368) recapitulates his main message: « My pursuit of what I have called the Spearman hypothesis ... represents an effort [to bring] the Black-White difference into the whole nomothetic network of the g construct ».

2. PROBLEMS OF DEFINITION: LEVEL I AND LEVEL II OF SPEARMAN'S HYPOTHESIS

Although all these authors agree with Jensen on the profound implications of Spearman's hypothesis, they would hardly agree on its definition, because Jensen – who coined the term – defined it with several conflicting meanings:

Definition 1 of 'Spearman's hypothesis' was cited above from the Abstract of Jensen's (1985) article. Definition 2 appears on p. 194: « Spearman ... noticed that the mean difference was most marked in just those [tests] which are known to be most saturated with g' ... Since Spearman's observation was based on a rather limited and unreplicated set of data, it seems best to regard it ... as a hypothesis ». Definition 3 merges on p. 198: « The strong form of the hypothesis holds that the agnitude of the Black-White differences ... are directly related to the est's g loadings, because Black and White populations differ only on g nd no other cognitive factors". Definition 4 states « The weak form of he hypothesis holds that the Black-White difference in various mental ests is predominantly a difference in g, although the populations also iffer, but to a much lesser degree, in certain other ability factors besides g. » (p. 198).

These vague circumlocutions contain at least two technically distinct interpretations:

(i) Spearman hypothesis, Level I interpretation: the mean difference vector correlates positively with the regression weights of the PC1 of the *pooled covariance matrix*.

(ii) Spearman hypothesis, Level II interpretation: the mean difference vector correlates positively with the regression weights of the PC1s of both *within sample covariance matrices*.

Jensen used both interpretations interchangeably. In (Jensen, 1980), he explicitly referred to the Level I version: « Probably the most compelling assemblage of evidence for the Spearman hypothesis ... is the massive data of the General Aptitude Test of the US Employment Service » (p. 549). « The correlations were not computed separately for Black and White samples but are based on predominantly White samples » (Jensen, 1985, p. 216). Hence this compelling evidence, a positive correlation (r = .71) between the first principal component of the *pooled* correlation matrix and the mean difference vector, confirmed the Level I version, but not the Level II version of his definitions.

In Jensen (1985, p. 199), he added several 'Methodological desiderata', one of which was that « Ideally ... factor analysis should be performed in the two population groups separately » – now appealing to the Level II version.

Strictly speaking, a 'desideratum' does not become part of a definition as long as it is not explicitly included in it. While it is a desideratum that husbands, ideally, should be faithful, faithfulness is not part of the definition of 'husband': unfaithful husbands are still husbands, albeit not ideal ones. However, since most of Jensen's data, though not his definitions, actually do point to the stronger Level II interpretation (Shockley, 1987), these two technically distinct versions of Jensen's shifting definitions of the term 'Spearman's hypothesis' should be clearly distinguished.

The main point of the present paper is to show that the positive correlations predicted by Spearman's hypothesis can be accounted for as psychometric artefacts (Note 1) under both interpretations. They, therefore,

(a) have nothing to do with « the whole nomothetic network of the g construct », because they also arise with data which contain no g at all, e.g., variables which measure the number of toys a child has and how often it plays with them (section 4),

(b) and they have nothing to do with Whites and Blacks and the nation's economic welfare, because they also arise with any set of moderately correlated random numbers which contain no g (section 6).

3. INADEQUACY OF MATHEMATICAL MACHINATIONS

In (Schönemann, 1985) I presented an algebraic argument which shows that, under the Level I interpretation, « the predicted correlation ... is a psychometric artefact that arises with any data as long as the covariance matrices are equal and the mean vectors are sufficiently different » (p. 241). I illustrated this with two computer simulations (see Table 1, loc. cit.), and further showed by geometric argument (summarized in sec. 5, below) that the equal covariance assumption is unnecessary.

In response, Shockley (1987 – not Jensen, 1985) challenged the Level I version as too narrow to do justice to all of Jensen's data. This valid criticism prompted me to extend my geometric argument to Level II by adjoining a multinormality and a positivity assumption (cf. sections 5 and 8, below).

Jensen (1987) views all such algebraic demonstrations and simulations with disdain: « Why use ficticious examples? If Schönemann really has a valid argument, why not use it to show, for example, that g, or the largest common factor extracted from different batteries of cognitive tests, is not highly similar across the different batteries ...? » (cf. Note 2). « ... The reason Schönemann cannot do this is simply that individual differences and the mean differences between populations on a great variety of cognitive tests do not depend in the least on the mathematical machinations demonstrated in his ficticious examples » (Jensen, 1987, p. 387).

4. THE GENERAL TOY FACTOR OF THE HEAD START DATA

To parry this challenge, real data are called for. Such data are ntained in a report of the Westinghouse Learning Corporation (1969, ote 3), *The impact of head start*. Among many other variables, these ta contain the following eight measures which form 'positive mani- lds' (Note 4) in several subsamples of Black and White children:

ead start toy variables and subsamples

ariable 1: Number of toys that child has which could be used in playing school
(col. 23. 1: 0, 2: 1-2, 3: 3-5, 4: 6-9, 5: 10 or more)

'ariable 2: Number of books child has to read
(col. 24. 1: 0, 2: 1-2, 3: 3-5, 4: 6-9, 5: 10 or more)

'ariable 3: How often child reads by himself at home
(col. 25. 1: seldom or never, 2: sometimes, 3: often, 4: regularly, 5: extremely often)

'ariable 4: How often respondent reads with child
(col. 26. 1: seldom or never, 2: sometimes, 3: often, 4: regularly, 5: extremely often)

/ariable 5: Length of time child reads or was read to on the day before the interview
(col. 27. 1: not at all, 2: up to 15 minutes, 3: 15-30 minutes, 4: 30 minutes-1 hour, 5: more than 1 hour)

/ariable 6: Number of games child has
(col. 28. 1: none, 2: 1-2, 3: 3-5, 4: 6-9, 5: 10 or more)

/ariable 7: How often child plays with games
(col. 29. 1: seldom or never, 2: at least once a week, 3: several times a week, 4: at least once a day, 5: at least several times a day)

Variable 8: How often respondent plays games with child
(col. 30. 1: seldom or never, 2: at least once a week, 3: several times a week, 4: at least once a day, 5: at least several times a day)

Stratifying variable: Race
(col. 185. 1: White, 2: Black).

The analysis had to be limited to the Summer Programs data of the Head Start program, because the sample sizes for the Full Year Programs are too small. The Summer Program subsamples are described in Table 1.

These data were analyzed by eight subsamples (in pairs A White and A Black, and B White and B Black, respectively) and also by six total samples. A detailed illustration is set out in Table 2. The eight Toy variables evidently form a well-defined 'positive manifold' (that is, are positively correlated throughout, cf. Note 4).

Collinearity measures computed on the toy data

For each subsample, the following statistics were computed:
(a) means, variances, and correlations of the eight Toy variables,
(b) the mean Black/White difference vector was standardized by dividing out the pooled standard deviations (see, e.g., the formula in Jensen, 1985, p. 199, except that the weighting was done in terms of df instead of sample sizes), resulting in the standardized mean difference vector d,
(c) the regression weight vectors of the first (largest) principal components (PC's) a_w of the White correlation matrix and a_b of the Black correlation matrix and the percent of variance each PC accounted for, v_w and v_b,
(d) the correlations r_{wd}, r_{bd}, between d and the two regression weight vectors a_w, a_b, which from now on will be called 'Spearman hypothesis correlations',
(e) the cosines ('congruence coefficients') c_{wd}, c_{bd} between the two regression weight vectors and the mean difference vector as an alternative collinearity index, and
(f) the cosine c_{wb} between a_w and a_b.

Results of the analysis of the head start toys data

The results for the 8 smaller subsamples, A, B, and also for the 6 larger total subsamples by year are given in Table 3. As can be seen, most Spearman hypothesis correlations r_{wd}, r_{bd} predicted by the Level II interpretations are positive. For the smaller samples, they are more variable than for the larger samples. However, their average is positive for both sets: .47 for the 6 large White samples and .39 for the 6 large Black samples. For the larger samples, all 12 correlations are positive.

able 1
ubsamples of the summer headstart toy data

Year	Treatment	Subsample	Race	Code	Sample size
1st	head start	A	White	s1haw	171
		B		s1hbw	107
		total		s1htw	278
		A	Black	s1hab	72
		B		s1hbb	121
		total		s1htb	193
	control	A	White	s1caw	191
		B		s1cbw	128
		total		s1ctw	319
		A	Black	s1cab	64
		B		s1cbb	101
		total		s1ctb	165
2nd	head start	A	White	s2haw	159
		B		s2hbw	115
		total		s2htw	274
		A	Black	s2hab	94
		B		s2hbb	92
		total		s2htb	186
	control	A	White	s2caw	177
		B		s2cbw	129
		total		s2ctw	306
		A	Black	s2cab	77
		B		s2cbb	84
		total		s2ctb	161
3rd	head start	total	White	s3hw	199
		total	Black	s3hb	159
	control	total	White	s3cw	229
		total	Black	s3cb	130

Table 2
Illustration of the head start toy data computations

Subsamples: Summer Program, 3rd year, Controls (s3cw, s3cb)
Sample sizes: $N_w = 229$, $N_b = 130$

Variable	White mean	White s.d.	Black mean	Black s.d.	d	1st PCs a_w	1st PCs a_b
1. Number of toys	3.65	1.29	2.85	1.19	.63	.62	.70
2. Number of books	3.76	1.39	2.87	1.50	.62	.75	.62
3. Child reads	2.45	1.21	2.27	1.23	.12	.61	.61
4. Respondent reads	1.86	.94	1.68	.90	.18	.56	.62
5. Read day before	1.60	1.07	1.36	.83	.25	.40	.65
6. Number of games	2.89	1.21	2.25	.96	.57	.72	.71
7. Child plays	2.27	1.04	2.35	1.27	-.07	.65	.62
8. Respondent plays	1.60	.80	1.46	.79	.18	.64	.67

Correlations (White above, Black below diagonal)

	1	2	3	4	5	6	7	8
1. Number of toys	1.00	.50	.17	.23	.04	.47	.24	.32
2. Number of books	.52	1.00	.43	.29	.30	.53	.27	.28
3. Child reads	.31	.33	1.00	.29	.52	.22	.25	.27
4. Respondent reads	.34	.26	.40	1.00	.07	.23	.28	.43
5. Read day before	.35	.29	.46	.45	1.00	.12	.18	.07
6. Number of games	.45	.45	.24	.26	.27	1.00	.51	.31
7. Child plays	.29	.17	.29	.19	.24	.55	1.00	.46
8. Respondent plays	.36	.24	.24	.37	.37	.42	.49	1.00

Summary statistics	White		Black
Percent of variance accounted for by 1st PC	.39		.43
Spearman hypothesis correlations	.37		.60
Cosines	.81		.80
Cosine between both within-sample PCs		.99	

Table 3
Results of the analysis of the head start toy data

subs.		N_w	N_b	v_w	v_b	r_{wd}	r_{bd}	c_{wd}	c_{bd}	c_{wb}
Small samples										
s1caw s1cab	A	191	64	.38	.46	*.13*	*.38*	.75	.77	1.00
s1cbw s1cbb	B	128	101	.41	.41	*.90*	*.54*	.79	.73	.99
s1haw s1hab	A	171	72	.42	.46	*-.23*	*-.41*	.00	-.02	1.00
s1hbw s1hbb	B	107	21	.39	.46	*.25*	*.17*	.44	.43	.99
s2caw s2cab	A	177	77	.43	.40	*.36*	*.20*	-.31	-.32	1.00
s2cbw s2cbb	B	129	84	.35	.42	*.45*	*.35*	.83	.82	.99
s2haw s2hab	A	159	94	.39	.42	*.53*	*.31*	-.05	.00	.97
s2hbw s2hbb	B	115	92	.40	.43	*.27*	*.26*	.70	.69	.99
sums and means		1,177	605	.37	.43	*.34*	*.19*	.39	.38	.99
Large samples										
s1ctw s1ctb		319	165	.39	.43	*.72*	*.46*	.82	.80	.99
s1htw s1htb		278	193	.40	.46	*.22*	*.06*	.36	.34	1.00
s2ctw s2ctb		306	61	.40	.41	*.49*	*.31*	.59	.58	.99
s2htw s2htb		274	186	.40	.42	*.32*	*.40*	.39	.42	.99
s3cw s3cb		229	130	.39	.43	*.37*	*.60*	.81	.80	.99
s3hw s3hb		199	159	.38	.43	*.69*	*.50*	.11	.05	.99
sums and means		1,605	994	.39	.43	*.47*	*.39*	.51	.45	.99

Note: Spearman correlations between standardized mean difference vector d and within sample first principal axes (predicted by Spearman's hypothesis, Level 2 interpretations) in italics. For column headings see text. For sample code see Table 1.

Essentially the same picture emerges if one measures the collinearity between the standardized Black/White difference vector and the two within-sample principal components in terms of cosines which are actually more appropriate collinearity measures. The average cosines are .51 and .45 for the White and Black samples, respectively.

Finally, the cosines between the regression weight vectors of the White and Black subsamples average .992 for the head start toy data.

According to Gordon (1987), « By the standards usually applied to congruence coefficients ... this indicate[s] that g factors and Black-White factors definitely represent the same construct » (p. 127, cf. Note 2).

In the present case, this construct is evidently a General Toy factor, objectively measured and determined.

Verification that the toys PC1 is not *g*

A reviewer cautioned that a « weakness of the toy analyses is that followers of Spearman's hypothesis will simply argue that the toy and reading questions are legitimate correlates of intelligence ».

The fact that height and weight correlate moderately (r = .6) does not mean that height and weight are the same thing. Nevertheless, this criticism would acquire some force if the Toys PC1 were to correlate strongly with an 'intelligence test' PC1. It does not:

The PC1 of the Toy variables was correlated with the PC1 of the six Metropolitan Readiness Tests (MRT), « a group-administered test with six subscores and a total readiness score [which] was used to assess the readiness for academic learning in grade 1. This test does not require the ability to read » (Westinghouse Learning Corporation, 1969, p. 135). The sample consists of 300 first graders of the Summer Program pooled across ethnic groups. As can be seen from Table 4, both the Toy variables and the cognitive variables again form 'positive manifolds'. The PC1 of the Toy variables has variance 3.27, and the PC1 of the MRT has variance 3.53, the covariance between both PC1s is .91. The implied low correlation, .27, clearly indicates that two PC1's of the 'intelligence tests' and of the toy variables are different.

The fact that the 'Toy factor' is not the 'g' of the MRT could also have been inferred, more directly, by inspecting the correlations. The within-battery correlations are considerably higher on average than the between-battery correlations, which is why the correlation of the two PC1s is so low (Note 5).

Table 4
Correlations of the head start toy variables and the metropolitan reading readines tests (MRT)

A. Toys, R_{11}	1	2	3	4	5	6	7	8
1. Number of toys	1.00	.51	.33	.37	.27	.48	.32	.32
2. Number of books	.51	1.00	.41	.48	.33	.40	.27	.24
3. Child reads	.33	.41	1.00	.35	.29	.15	.14	.19
4. Respondent reads	.37	.48	.35	1.00	.42	.18	.14	.32
5. Read day before	.27	.33	.29	.42	1.00	.10	.08	.20
6. Number of games	.48	.40	.15	.18	.10	1.00	.60	.40
7. Child plays	.32	.27	.14	.14	.08	.60	1.00	.62
8. Respondent plays	.32	.24	.19	.32	.20	.40	.62	1.00
1st eigenvector, s_1':	.73	.73	.54	.63	.49	.67	.64	.65
Largest eigenvalue:	3.27							

B. MRT, R_{22}	9	10	11	12	13	14
9. Word meaning	1.00	.59	.43	.36	.53	.23
10. Listening	.59	1.00	.55	.40	.63	.46
11. Matching	.43	.55	1.00	.49	.63	.57
12. Alphabet	.36	.40	.49	1.00	.62	.45
13. Number	.53	.63	.63	.62	1.00	.58
14. Copying	.23	.46	.57	.45	.58	1.00
1st eigenvector, s_2':	36	.42	.43	.38	.47	.38
Largest eigenvalue:	3.53					

C. Between-set correlations, R_{21}

	1	2	3	4	5	6	7	8
9.	.11	.10	-.07	.10	-.04	.16	-.01	-.03
10.	.15	.19	.06	.04	.03	.20	.05	.01
11.	.24	.17	.13	.07	.00	.27	.11	.08
12.	.21	.22	.13	.17	.09	.14	.05	.13
13.	.30	.26	.14	.09	.02	.30	.11	.12
14.	.34	.22	.14	.09	.07	.25	.09	.10

$\mathrm{cor}(PC1_{TOYS}, PC1_{MRT}) = s_1'R_{12}s_2/(s_1'R_{11}s_1 s_2'R_{22}s_2)^{1/2} = .27.$

Sample: First Year Summer Program, A, all ethnic groups. N = 300.

5. A GEOMETRIC EXPLANATION OF THE SPEARMAN HYPOTHESIS ARTEFACT

So as not to reiterate the algebraic argument in Schönemann (1985) to explain the weaker, Level I, interpretation of 'Spearman's hypothesis', an intuitively more direct geometric argument will be given here which extends to the stronger Level II version if two additional assumptions are adjoined:

Level I (Figure 1A)

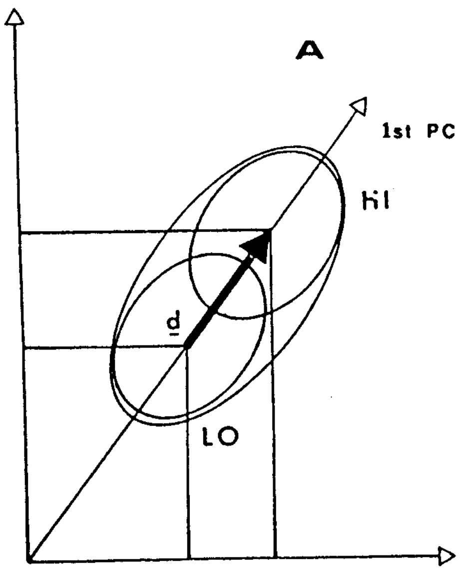
If we have two groups, HI and LO, and pull them apart so that the length of the mean difference vector d := y1-y2 increases, then the pooled group will contain a dominant first eigenvector (the main axis of the pooled group), which will become increasingly more parallel to d. This will occur regardless whether both subgroups form 'positive manifolds' or not.

Level II (Figure 1B)

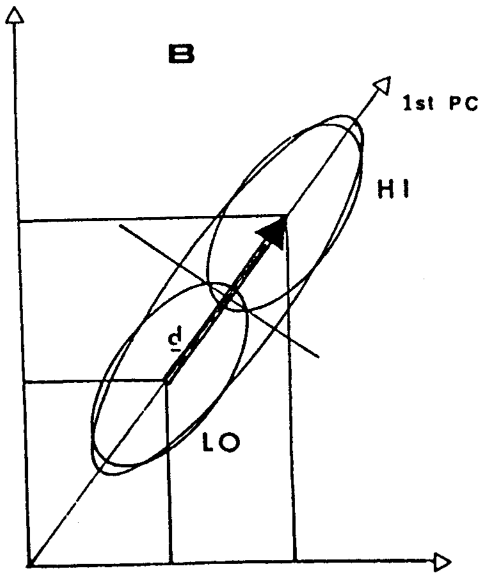
To extend this reasoning to Level II, one needs (a) a strong 'positive manifold' in the pooled population, and (b) approximate multinormality in the pooled distribution (i.e., that the equidensity contours are ellipsoidal), as in Figure 1B. Then any roughly even split into a HI and LO group produces two attenuated within-covariance matrices. As long as they remain positive, their principal axes will be approximately collinear with the principal axes of the pooled population, so that the PC1s of all three populations will correlate highly with the mean difference vector *d*.

Loosely, if one cuts a banana into two halves (and both remain elongated), then the main axes of both halves will be roughly parallel to the main axis of the whole banana, and all three will be roughly parallel to the line segment connecting the centroids (points of gravity) of both halves.

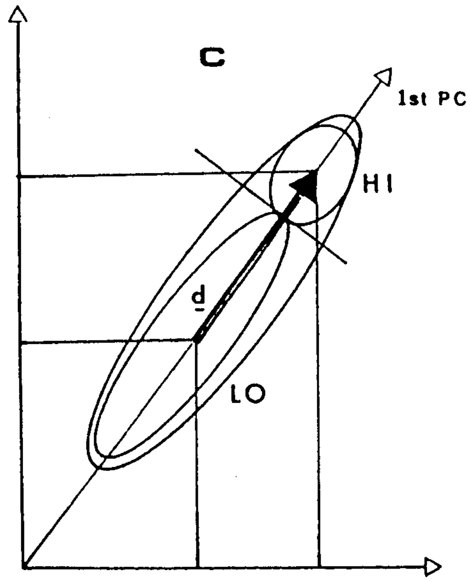
Figure 1
A. Spearman's hypothesis, Level I.
If we start with a HI and a LO sample, and pull them apart, thus increasing the length of the mean difference vector d (heavy arrow), then d will be approximately collinear with the first principal component PC1 of the pooled sample (main axis of larger enveloping ellipse).



B. Spearman's hypothesis, Level II.
If we start with a pooled sample which contains a dominant PC1 (dominant main axis of larger, enveloping ellipse) and cut it into two subsamples HI and LO of roughly equal size (smaller ellipses), then the mean difference vector d (heavy arrow) will not only be parallel to the main axis of the larger ellipse as in A, but it will also be parallel to the main axes of the two smaller ellipses.



C. Spearman's hypothesis, Level II, Unequal splits prediction.
For uneven splits, the main axis of the larger sub sample (LO in the picture) will be more nearly parallel to the first PC of the pooled sample, and thus to d, than the main axis of the ellipse of the smaller sample whose eccentricity is smaller.



6. UNEQUAL SPLITS PREDICTION

This intuitive geometry suggests a strong prediction which is easily checked on real data.

Unequal splits prediction:

> If the sample size of one group is much larger than the sample size of the other, then the correlation between the mean difference vector and the first principal axes will be larger for the larger sample.

There are two reasons for this: (a) the joint distribution of the smaller sample will be less eccentric (more nearly circular, cf. Figure 1C and last line in Table 3), so that its PC1 is less well defined. In addition, (b), the estimate of the within-sample covariance matrix for the smaller sample will be more fallible. More generally, all these predictions rely on reasonably large sample sizes to ensure that the ellipsoidal equidensity contour implied by the pooled group 'positive manifold' is well defined.

Verification of the unequal splits prediction in the toys data

The two total subsamples for the Whites were pooled over treatments so as to arrive at White samples with substantial sample sizes of $N = 400$. These larger samples were then analyzed with each of the two smaller treatment samples (headstart and control) for the Blacks. For these uneven Black/White splits the Spearman correlations were:

1st year $N_w = 400 > 165 = N_b$, $r_w = .56 > .36 = r_b$
$N_w = 400 > 193 = N_b$, $r_w = .27 > .08 = r_b$
2nd year $N_w = 400 > 161 = N_b$, $r_w = .56 > .30 = r_b$
$N_w = 400 > 186 = N_b$, $r_w = .60 > .45 = r_b$

However, for the
3rd year $N_w = 400 > 130 = N_b$, $r_w = .36 < .56 = r_b$
$N_w = 400 > 159 = N_b$, $r_w = .51 < .54 = r_b$

this effect breaks down, presumably because the sample sizes of the two Black samples are still too small to ensure reliable correlation estimates.

Verification of the unequal splits prediction in Jensen's data

For Jensen's (1985) reanalysis of the 'intelligence test' batteries (his Table 3), the results were as follows:

Jensen & Reynolds: $N_w = 1{,}868 > 305 = N_b$, $r_w = .73 > .54 = r_b$.
N'l Longitudinal: $N_w = 12{,}275 > 1{,}938 = N_b$, $r_w = .78 > .68 = r_b$.
Nichols: $N_w = 1{,}940 > 1{,}460 = N_b$, $r_w = .75 > .71 = r_b$.
Dept. of Defense: $N_w = 5{,}533 > 2{,}298 = N_b$, $r_w = .39 > .29 = r_b$.
Kaufman & Kaufman: $N_w = 813 > 486 = N_b$, $r_w = .56 > .49 = r_b$.

The only study which seems to violate this rule is Hennesey and Merrifield, but it involves partial correlations. The remaining studies either involve relatively small sample sizes or approximately even Black/White splits.

Verification of the unequal splits prediction for random data which contain no *g*

In Schönemann (1985), the results of a computer simulation with random data were given to illustrate the artefact under the Level I interpretation of Jensen's definition of Spearman's hypothesis.

To study the effect of varying sample sizes at Level II, 200 simulations were run with random data, as follows:

Generation of 'positive manifolds' devoid of *g*

A p × p 'positive manifold' for the pooled sample was constructed as $C = TT' + sI$, where the elements t_{ji} of T were uniformly [0,1] distributed random numbers, I is the p × p identity matrix and s is a scalar > 0. This ensures that C is positive definite for all $s > 0$. In the simulation, s was set to 1. On normalizing C, one obtains a randomly generated correlation matrix R with an average correlation of about .6 (for $s = 1$) which does not contain a general factor. This matrix was treated as a pooled correlation matrix R and factored into $R = AA'$, where A is nonsingular.

The right hand factor, A', was applied to a matrix of independent normally distributed random deviates Z of order $(N_1 + N_2) \times p$ to arrive at a pooled score matrix $Y = ZA'$ with approximate correlation matrix $AA' = R$. Y was split into two subsamples on the basis of the total scores (Note 6): The N_1 'subjects' with the highest total scores were assigned to Y_1 (HI) and the remaining N_2 'subjects' to Y_2 (LO). After extracting (the weights of) the first principal components a_p of the pooled correlation matrix R and the first principal components a_1, a_2 of the two within correlation matrices R_1 and R_2, the Spearman correlations r_{pd} (predicted by the Level I interpretation of 'Spearman's hypothesis'), and r_{1d}, r_{2d} (predicted by the Level II interpretation), and the cosines c_{pd}, c_{1d}, c_{2d}, and c_{12}, were computed. These statistics were then averaged over 25 replications each for various combinations of two subsample sizes N_1, N_2 and number of variables p.

Results

The results of this random simulation are summarized in Table 5. Together with the percentages of variance accounted for by the first principal components in the pooled sample (vp) and the two subsamples (v_1 and v_2), the correlations and cosines between the difference vectors *d*, and the regression weights of the first principal components of the two subsamples, are given. Over all 200 replications, the average Spearman hypothesis correlation was .593 (vs. Jensen's .59). As long as the within sample sizes were reasonably large (> 200), this correlation was never negative. The cosines between the first eigenvectors of all three (two within and one pooled) samples were highly positive and roughly of the same order as those reported in Jensen (1980, 1985).

Table 5
Computer simulation of the within sample version (II) of Spearman's hypothesis

N_1	N_2	p	n	v_p	v_1	v_2	r_{1d}	c_{1d}	r_{1p}	r_{2d}	c_{2d}	r_{2p}	c_{12}
100	100	10	25	.615	.311	.297	**.569**	.979	.678	**.594**	.942	.719	.934
150	150	5	25	.587	.299	.309	**.612**	.753	.713	**.486**	.746	.643	.707
150	150	10	25	.615	.306	.302	**.602**	.979	.767	**.621**	.981	.783	.979
150	150	15	25	.646	.327	.324	**.531**	.991	.748	**.552**	.992	.755	.989
200	200	5	25	.572	.305	.298	**.721**	.630	.743	**.629**	.700	.755	.736
200	200	10	25	.621	.304	.310	**.716**	.980	.816	**.685**	.981	.839	.981
200	200	15	25	.649	.336	.330	**.661**	.992	.828	**.679**	.992	.807	.991
100	300	10	25	.618	.192	.486	**.103**	.139	.174	**.736**	.998	.929	.144
Means over 175 simulations, excluding uneven splits (last row)				.615	.313	.310	**.630**	.901	.756	**.607**	.905	.757	.902
Means over all 200 simulations				.615	.298	.332	**.564**	.805	.683	**.623**	.916	.779	.808

N_1:= sample size for HI group N_2:= sample size for LO group
p:= number of variables n:= number of simulations

Percent of variance accounted for by 1st principal component:
v_p:= in pooled population, v_1:= in HI sample, v2:= in LO sample

(Spearman Hypothesis) Correlations of 1st eigenvectors with mean difference vector d: r_{1d}:= in HI sample r_{2d}:= in LO sample

Cosines of 1st eigenvector with mean difference vector d:
c_{1d}:= in HI sample c_{d2}:= in LO sample

Correlations of 1st eigenvectors within samples with 1st eigenvector in population: r_{1p}:= in HI sample r_{2p}:= in LO sample

Cosines of both within-sample eigenvectors: c_{12}

The values in each row of the Table are averages over 25 simulations per (N1, N2, p) parameter configuration. The correlations predicted by Spearman's hypothesis are in bold type.

(For a simulation under the Level I interpretation of Spearman's hypothesis, see Schönemann, 1985.)

For unequal splits (last row in Table 5), the larger sample has the larger Spearman hypothesis correlation, just as was the case for Jensen's data and the Head Start Toy data.

7. RELATED QUESTIONS OF THE REALITY OF THE SPEARMAN HYPOTHESIS EFFECT

Humphreys (1985a) has discussed some data which in some respects seem to corroborate the foregoing analysis, while in other respects they seem to conflict with it. He analyzed 54 cognitive tests in large subsamples of the Project Talent data, computing the mean difference vector of two High and Low SES groups for the Whites (a 5/1 split), the mean difference vector of the Blacks versus Low SES Whites (a 1/1 split), and the total White-Black mean difference vector (6/1 split) with the loadings based on the school means of the same students. He found the highest Spearman hypothesis correlation (about .8) for the High White/Low White contrast, the smallest for the Low White/Black contrast (.2) and an intermediate value (.5) for the White/Black contrast.

These empirical results corroborate our conclusion from the Toy data analyses that positive Spearman hypothesis correlations do not warrant any inferences about the nature of *g* and Black-White differences: « It is highly probable that [Jensen] would find equal or stronger support in the same tests for socio-economic differences in the White population » (Humphreys, 1985a, p. 283).

On the other hand, they appear to conflict with the prediction that these correlations should be largest for the largest group if both 'positive manifolds' are sufficiently well defined. However, Humphreys' results are not easy to evaluate because (a) he did not report his data in sufficient detail to be able to check this, and (b), as he himself points out, his analyses deviated from the setup required to test the Level II versions of Spearman's hypothesis stringently. In particular, he « did not consider separate analyses essential. My loadings were for boys who represented all races and ethnic groups in the 10th grade in 1960. Their loadings happen to be in the more convenient column of the published table ... » (Humphreys, 1985b, p. 292). Thus, in effect, he dealt with the Level I version of Spearman's hypothesis which permits no predictions about the relative magnitude of the Spearman hypothesis correlations as a function of the location of the dividing hyperplane.

In contrast, an early study by Nagoshi, Johnson, Defries, Wilson, and Vandenberg (1984), based on 15 tests of specific cognitive abilities which had been administered to 1816 sets of parents and 2949 teen-aged or older off-spring of Japanese, Chinese, and European ancestry, did find within-group (i.e., Level II) Spearman hypothesis correlations both across ethnicity and generation. The ethnicity correlations were smaller than those reported by Jensen.

In agreement with the overall message delivered here, these authors also question the empirical reality of Jensen's Spearman hypothesis correlations, though on less formal grounds:

« Because a group difference on *g* requires group differences on tests which load on *g*, an observed group difference in general mental ability may necessarily result in a correlation between group differences on individual tests and their *g*-loadings » (p. 751; see also Wilson, 1985).

8. RECENT THEORETICAL DEVELOPMENTS

Since the first draft of this paper was written (and submitted to *The Behavioral and Brain Sciences*, in November 1986), further progress has been made on the formal aspects of the Level II version of Spearman's hypothesis. In Fall 1987, L. Guttman gave me a copy of a manuscript which has since appeared posthumously as a target article in the *Multivariate Behavioral Research* (Guttman, 1992). It contains a trenchant critique of many of Jensen's claims about Spearman correlations, positive manifolds, *g*, reaction time and Black/White differences (see also Horn & Goldsmith, 1981, for an earlier comprehensive critique of some of Jensen's claims).

In his paper, Guttman proved that « actual proportionality must hold between the loadings and the standardized [mean] differences [if] Spearman's tetrad condition holds for each of the subpopulations as well as for the total population ».

However, as he also pointed out, this assumption is, of course, never satisfied empirically, notwithstanding Jensen's constant references to *g*, which strictly speaking require that it be satisfied: « Any reader of these lines can himself easily disprove '*g*' by looking at almost any mental test correlation matrix at his disposal and checking for proportionality » (p. 13).

Since Jensen in effect has abandoned Spearman's factor model by working with principal components instead of factors, Guttman's result is not directly applicable to the Level II version of Spearman's hypothesis. However, the following result can be proved which does not require that the Spearman model holds in all three populations, because it involves principal components rather than factors:

Theorem: If the range Re^p of a p-variate normal random vector y $N_p(\varnothing, \Sigma)$, where Σ is positive with $diag(\Sigma) = diag(I)$, is partitioned into a High set (H) and a Low set (L) by the plane orthogonal to the PC1, and containing the origin, and both within-covariance matrices Σ_H, Σ_L remain positive, then (a) the mean difference vector $d := E(y|H)-E(y|L)$ will be collinear with the PC1 of the pooled Σ, and (b) d will also be collinear with both within PC1s of Σ_L, Σ_H.

For the proof, see (Schönemann, 1989, 1992). The original formulations referred to a cutting plane orthogonal to the first centroid instead of the PC1, which coincide only if $p = 2$. In higher dimensional spaces, they will be close but not strictly collinear. However, both the accompanying graph (Figure 1) and the proof left little doubt about the intended location of the cutting plane (cf. Note 6).

Concretely, under the stated assumptions, the *cosines* between the mean difference vector and the largest eigenvectors and the two within-sample *covariance* matrices are not just positive in general but, except for sampling error, are unity in all multinormal distributions with positive within-covariance matrices. This fact is obscured when the analyses are based on correlations instead of covariances and when collinearity is measured in terms of correlations instead of cosines.

The above theorem is more general than Guttman's because it does not rest on the unrealistic assumption that the factor model holds in any of the three populations. Rather, it is stated in terms of principal components, which always exist. Note, however, that the proof requires perfectly even splits.

In practice, of course, formal assumptions required in proofs can be expected to hold only approximately, not exactly. Therefore, in practice, the conclusions can also be expected to hold only approximately, not exactly. For more on this, see Note 6.

9. DISCUSSION

It thus emerges that positive Spearman hypothesis correlations can be discounted as artefacts under *both* interpretations of Spearman's hypothesis: They are a tautological consequence of two well-known facts, (a) that the subtests of IQ-tests, essentially by construction, correlate positively, and (b) that Blacks – for whatever reasons – at this point in time, score lower than Whites on these tests.

In particular, they tell us nothing about 'the nature of *g*' (because the random matrices in section 5 did not contain any *g*'s). They do not warrant any of the cosmic and sometimes comic extrapolations some have derived from them.

As the quotes in section 1 attest, Shockley (1971) and Jensen (1985) are not alone in seeming to delight in stigmatizing whole ethnic groups and all their descendents on the basis of flimsy evidence and flawed reasoning. For a more recent illustration, see Herrnstein and Murray (1994, especially p. 726), which further publicized Jensen's presumed discovery while carefully pruning away any contrary evidence.

Such excesses would be unthinkable if the reviewing system functioned as it is supposed to. That it does not has been pointed out before (see, e.g., Hearnshaw, 1979; Hirsch, 1981; Peters & Ceci, 1982). Yet nowhere is the potential social cost of unchecked charlatanism higher than in the mental test area, which has deep historical roots in institutional racism (cf., e.g., Gould, 1981; Moss, 1985; Robitscher, 1973).

The non-chalant dismissals of valid criticisms of Jensen's bold conjectures by various editors of the American psychological establishment is especially worrisome in view of the social significance of Jensen's claims. Surely they should at least have noticed that we are not dealing with a trivial topic. Of all the evidence available, Jensen's positive Spearman hypothesis correlations seemed to present the strongest case so far for the existence of a general mental ability. If they had not been artefacts, they would have been the discovery of the century: For the first time we would have an independent corroboration of the psychometric *g* construct which, up to that point, was supported only by internal validation, and by the specious argument that '*g*'s' across different batteries must be similar because the regression weights are similar (Note 2).

On closer inspection, it turns out that all attempts to reify Spearman's *g*, either internally by appeal to the factor model, or externally by appeal to predictive validities, heritability ratios, or Spearman correlations, were failures.

Claims of internal, psychometric validations have collapsed because (a) Spearman's factor model virtually never fits any data – one invariably needs more than one common factor to describe conventional 'intelligence tests' – and (b) because this model has a built-in indeterminacy which defeats its very purpose to provide an *objective definition* of *g* (Maraun, 1996; Mulaik, 1986; Schönemann, 1971, 1978, 1981a, 1981b, 1987, 1996a, 1996b; Schönemann & Wang, 1972; Steiger & Schönemann, 1976).

Claims of external validation in terms of predictive validities have failed because omnibus tests, the commercial 'intelligence tests', and the closely related 'scholastic aptitude' tests, do a remarkably poor job of predicting even criteria they were specifically designed to predict, such as college GPA. For example, Schrader (1971, p. 127) presents validity coefficients of the SAT and highschool record. On p. 142 he lists the SAT validities for college freshman science grades. For samples with at least 200 subjects, the average is .12 for the SAT-V and .19 for the SAT-M. Humphreys (1968) computed the predictive validities of a composite of the American College Testing (ACT) program for GPAs across eight semesters. Starting from a respectable .48, they steadily decline as a function of the prediction interval to .16 for the 8'th semester GPA (p. 376).

Claims of external validation of *g* by appeal to high heritability ratios collapsed when it became clear that the assumptions underlying these ratios are not only implausible (as had been suspected before, e.g., Hirsch, 1981; Wahlsten, 1990) but are, in fact, refuted by most data (Schönemann, 1990; Schönemann & Schönemann, 1994). Moreover, it emerged that one of the most popular 'heritability estimates' is unsound (Schönemann, 1993, in press).

It should therefore come as no surprise that Jensen's widely hailed effort to resuscitate Spearman's *g* by appeal to Spearman's hypothesis was also a failure. Yet since its inception, Spearman's *g* has been invoked constantly to shore up otherwise unsubstantiated claims about the 'success' of psychometrics (e.g., Estes, 1992) and the presumed inferiority of various ethnic groups, (Brigham, 1923; Jensen, 1970; Pearson & Moul, 1925; Herrnstein & Murray, 1994. See, e.g., Chase, 1980; Gould, 1981; for historical reviews).

People still making such claims seem to be just as impervious to the lessons of history as they are to technical arguments, which they dismiss as 'sterile', 'nihilistic', 'sophistic diatribes', or 'mathematical machinations' (Jensen, 1987, p. 386). All this points to a more ominous meaning of 'bias in mental testing' than that addressed by Jensen in his (1980) book.

ACKNOWLEDGEMENTS

This paper was completed while the author was on leave from Purdue University. I am grateful for support by the Taiwanese Foundation for the Advancement of Outstanding Scholarship. I also wish to thank the faculty of the Psychology Department of the National Taiwan University for their gracious hospitality.

NOTES

Note 1. 'Artefact' is used here with the meaning stated in the *Nouveau Petit Larousse* (1972, p. 69): « Structure or phenomenon of arteficial or accidental origin encountered during an observation or experiment bearing on a natural phenomenon ». In particular, the fact that an artefact arises 'with real data' is not conclusive evidence against it: If the Spearman hypothesis correlations had not been observed for 'real data', there would be no need to discuss them.

Note 2. The myth that similarity of regression weights implies similarity of random variables is still widespread, although it has been repeatedly refuted. With the same logic one could conclude that two random variables must be the same if their variances are the same. The simplest refutation I know of goes like this: Suppose we have two batteries of p tests each, which have identical within-set covariance matrices, and all between-set correlations are exactly zero. Then the regression weights for both PC1s are identical, while the PC1s themselves correlate zero.

Note 3. I am grateful to Professor Cicirelli, Department of Psychological Sciences, Purdue University, for providing me with the data, and to Drs. T. Putnam and W. Whitson of the Purdue University Computing Center for transferring them onto diskettes.

Note 4. A reviewer of an earlier draft pointed out, correctly, that the widely used term 'positive manifold' is actually a misnomer. Thurstone, who coined the term, originally meant by it a pattern of non-negative loadings throughout

after rotation to simple structure (not just a positive correlation matrix, which is the current meaning). However, cursory reflection suggests that the original meaning of the term may conflate with the notion of simple structure once oblique rotation is allowed (for $m = 1$ the current meaning implies Thurstone's). So far as I know, no-one has ever investigated this question systematically, perhaps for this reason. In view of the wide currency of the looser sense, I have opted to retain it here, but enclosed it in quotation marks in deference to the reviewer.

Note 5. This is not atypical, and more generally, the reason why the predictive validities of most omnibus tests, such as the SAT, GRE, the Wechsler tests, and the Stanford-Binet, are usually quite low. It is much easier to construct internally consistent tests (by item analysis), than it is to construct tests which correlate highly with a broad range of external criteria. Once such tests are validated against outside criteria, one usually finds that the between-set correlations are low and often zero. A convincing example of this is given in Table 2 of (Holland and Richards, 1965), where the off-diagonal correlation matrix between ACT scores and Artistic, Scientific, and Social Accomplishment is effectively zero, while the within-battery correlations of the 4 ACT tests average in the 50's and those of the 18 Accomplishment variables in the 30's (p. 14). This pervasive finding is obscured when a few selected validity coefficients are reported for paltry sample sizes. Once sample sizes are substantial enough to be taken seriously, as in this case ($N = 3770$) or for the Schrader (1971) SAT data, then empty boasts of the practical utility of such omnibus test evaporate quickly. As these facts become more widely known, it becomes harder to keep one's faith in g, which may be why they are not more widely known.

Note 6. Assumptions:

(a) Multinormality
Some reviewers challenged the plausibility of the assumptions underlying the Level II interpretation of 'Spearman's hypothesis' (It is not needed for Level I). As noted in the text, it is in the nature of logical arguments that they start with assumptions in order to arrive at conclusions. Clearly, some assumptions are more critical than others for making a point, and some are more plausible than others for making it stick.
Among the former, I count multinormality, which is critical for Level II and, I believe, also eminently plausible. I confess to being somewhat baffled how anyone could question multinormality as a default assumption. Multinormality undergirds virtually all multivariate statistics practiced in the social sciences, including maximum likelihood factor analysis, LISREL, tetrachoric correlations, corrections for attenuation, most likelihood ratio tests, most classical test theory, etc. etc. Why should this ubiquitous and universally accepted premise all of a sudden become suspect when it is invoked to refute counterintuitive claims of social import?

Moreover, approximate multinormality of test data is easily verified simply by inspection, without any need to invoke painstaking generalizations of the CLT (which are available if needed). I have never seen any bivariate plots of (continuous) mental test data which strikingly violate binormality, including the Head Start data described in the text. Thus, it seems to me, the least one would have to do before one can mount a convincing case against the multinormality assumption is to first produce some empirical evidence which conflicts with it.

(b) Positivity of covariance matrices
This is indeed a substantive assumption which may or may not hold in all three populations. It is more likely to be satisfied in the pooled group than in the subgroups. The whole notion of a 'positive manifold', I believe, is a tautological consequence of test construction procedure: We normally do not include items in a battery if they correlate negatively with the other items. Inevitable overlap in item content yields ellipsoidal joint distributions of item sums (the tests) in the total population. If we restrict the range by partitioning it into a HI and a LO subgroup, then the correlations will go down, as is well known. Hence the smaller of both subgroups is more likely to yield smaller or even negative Spearman Hypothesis correlations.

(c) Composition of pooled group
The exact composition of the pooled distribution is not likely to materially affect the overall conclusion. Anyone who wishes is free to assume two partially overlapping ellipses which produce an elliptic envelope (as in Fig. 1c, which strikes me as the more likely scenario) or, if one prefers, two strictly disjoint subpopulations partitioned by a cutting plane orthogonal to the PC1 of the pooled group. To make the algebraic proof go through, a cutting plane containing the origin perpendicular to the PC1 had to be assumed. In the simulations, it was actually perpendicular to the first centroid. This minor difference did not materially affect the predicted outcome of strongly positive Spearman hypothesis correlations.
More generally, it does not make much difference whether we cut a banana with a finely honed razor into two elongated halves, or smash it near the middle with a sledge hammer: In either case the main axes of both halves – as long as they both remain elongated – will still be approximately parallel to the centroid difference vector, which is all that is needed to establish the arteficiality of the Spearman hypothesis correlations.

(d) Location of cutting hyperplane
The Uneven Splits prediction is, of course, the crux of the matter. As was shown, it is quite robust. At this stage, the problem is no longer to just explain the Spearman hypothesis correlations, but also to explain why their magnitude varies with the location of the cutting plane. Multinormality explains both.

REFERENCES

Angoff, W. H. (1971). *The College Board Admissions Testing Program: A technical report on research and development activities relating the Scholastic Aptitude Test and achievement tests*. New York: College Entrance Examination Board.

Brandt, C. (1985). Jensen's compromise with componentialism. *Behavioral and Brain Sciences, 8*, 222-223.

Brigham, C. (1923). *A study of American intelligence*. Princeton: Princeton University Press.

Carlson, J. S. (1985). The issue of g: Some relevant questions. *Behavioral and Brain Sciences, 8*, 224-225.

Cattell, R. B. (1985). Intelligence and g: An imaginative treatment of unimaginative data. *Behavioral and Brain Sciences, 8*, 227-228.

Chase, A. (1980). *The legacy of Malthus: The social cost of the new scientific racism*. Chicago, IL: University of Illinois Press.

Estes, W. K. E. (1992). Ability testing. Postscript on ability tests, testing, and public policy. *Cognitive Science, 5*, 278.

Gordon, R. A. (1987). Jensen's contribution concerning test bias: A contextual view. In S. Modgil & C. Modgil, C. (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 77-154). New York: Falmer Press.

Gould, S. J. (1981). *The mismeasure of man*. New York: Norton.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27*, 175-204.

Hearnshaw, L. S. (1979). *Cyril Burt psychologist*. New York: Vintage.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.

Hirsch, J. (1981). To 'unfrock the charlatans'. *Sage Race Relations Abstracts, 6*, 1-67.

Holland, J. L., & Richards, J. M. Jr. (1965). *Academic and Non-Academic Accomplishment: Correlated or Uncorrelated?* ACT Research Reports, No. 2. Iowa City: American College Testing Program.

Horn, J., & Goldsmith, H. (1981). Reader be cautious: A review of Bias in Mental Testing. *American Journal of Education, 89*, 305-329.

Humphreys, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology, 59*, 375-380.

Humphreys, L. G. (1985a). Race differences and the Spearman hypothesis. *Intelligence, 9*, 275-283.

Humphreys, L. G. (1985b). Attenuated hypothesis or attentuated test of hypothesis. *Intelligence, 9*, 291-295.

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

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A. R. (1985). The nature of Black-White differences on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8,* 193-263.
Jensen, A. R. (1987). Differential psychology: Towards consensus. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 353-399). New York: Falmer Press.
Jones, L. V. (1985). Golly g: Interpreting Spearman's general factor. *Behavioral and Brain Sciences, 8,* 233.
Kline, P. (1985). The nature of psychometric g. *Behavioral and Brain Sciences, 8,* 234.
Maraun, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research, 31,* 517-538.
Modgil, S., & Modgil, C. (Eds.) (1987). *Arthur Jensen: Consensus and controversy.* New York: Falmer Press.
Moss, G. L. (1985). *Toward the final solution: A history of European racism.* Madison, WI: University of Wisconsin Press.
Mulaik, S. A. (1986). Factor analysis and Psychometrika: Major developments. *Psychometrika, 51,* 23-34.
Nagoshi, C. T., Johnson, R. C., DeFries, J. C., Wilson, J. R., & Vandenberg, S. G. (1984). Group differences and first principal-component loadings in the Hawaii Family Study of Cognition: a test of the generality of 'Spearman's hypothesis'. *Personality and Individual Differences, 5,* 751-753.
Nettlebeck, T. (1985). What reaction times time. *Behavioral and Brain Sciences, 8,* 235.
Nichols, R. C. (1985). Intelligence and its biological substrate. *Behavioral and Brain Sciences, 8,* 236.
Pearson, K., & Moul, M. (1925). The problem of alien immigration into Great Britain, illustrated by an examination of Russian and Polish Jewish children. *Annals of Eugenics, 1,* 1-127.
Peters, D. P., & Ceci, S. J. (1982). Peer review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5,* 187-195.
Robitscher, J. (1973). *Eugenic sterilization.* Springfield, IL: Thomas.
Schönemann, P. H. (1971). The minimum correlation between equivalent sets of uncorrelated factors. *Psychometrika, 36,* 21-30. Erratum. *Psychometrika,* 1973, *38,* 149.
Schönemann, P. H. (1978). Some nontrivial implications of factor indeterminacy. Paper read at the Annual Meeting of the Society for Multivariate Experimental Psychology, Cincinnati, Ohio.
Schönemann, P. H. (1981a). Factorial definitions of intelligence: Dubious legacy of dogma in data analysis. In I. Borg (Ed.), *Multidimensional data representations: When and why.* Ann Arbor, MI: Mathesis Press.

Schönemann. P. H. (1981b). Power as a function of communality in factor analysis. *Bulletin of the Psychonomic Society, 17,* 57-60.
Schönemann, P. H. (1985). On artefical intelligence. *Behavioral and Brain Sciences, 8,* 241-242.
Schönemann, P. H. (1987). Jensen's g: Outmoded theories and unconquered frontiers. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 311-327). New York: Falmer Press.
Schönemann, P. H. (1989). Some new results of the Spearman Hypothesis artifact. *Bulletin of the Psychonomic Society, 27,* 463-464.
Schönemann, P. H. (1990). Environmental versus genetic variance component models for the identical twin data by Shields. *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology, 10,* 451-473.
Schönemann, P. H. (1992). Extension of Guttman's result from g to PC1. *Multivariate Behavioral Research, 27,* 219-224.
Schönemann, P. H. (1993). A note on Holzinger's heritability coefficient h^2. *Chinese Journal of Psychology, 35,* 59-65.
Schönemann, P. H. (1996a). The psychopathology of factor indeterminacy. *Multivariate Behavioral Research, 31,* 571-577.
Schönemann, P. H. (1996b). Syllogisms of factor indeterminacy. *Multivariate Behavioral Research, 31,* 651-654.
Schönemann, P. H. (in press). On models and muddles of heritability, *Genetica.*
Schönemann, P. H., & Schönemann, R. D. (1994). Environmental versus genetic models for Osborne's personality data on identical and fraternal twins. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 13,* 141-167.
Schönemann P. H., & Wang, M. M. (1972). Some new results on factor indeterminacy. *Psychometrika, 37,* 61-91.
Schrader, W. B. (1971). The predictive validity of College Board Admission Tests. In W. H. Angoff (Ed.), *The College Board Admissions Testing Program: A technical report on research and development activities relating the Scholastic Aptitude Test and achievement tests.* New York: College Entrance Examination Board.
Shockley, W. (1971). Models, mathematics, and the moral obligation to diagnose the origin of negro IQ deficits. *Review of Educational Research,* 369-377.
Shockley, W. (1987). Jensen's data on Spearman's hypothesis: No artifact. *Behavioral and Brain Sciences, 10,* 512.
Spearman, C. S. (1927). *The abilities of man.* New York: MacMillan.
Steiger, J. H., & Schönemann. P. H. (1976). A history of factor indeterminacy. In S. Shye (Ed.), Theory construction and data analysis in the social sciences (pp. 136-178). San Francisco, CA: Jossey Bass.
Westinghouse Learning Corporation (1969). *The impact of head start.* Akron, OH: Ohio University.

Wahlsten, D. (1990). Insensity of the analysis of variance to heredity-environment interactions. *Behavioral and Brain Sciences, 13,* 109-120.
Wilson, J. R. (1985). Jensen's support for Spearman's hypothesis is support for a circular argument. *Behavioral and Brain Sciences, 8,* 246.

Received 11 December, 1996
Accepted 3 May, 1997

CPC

Cahiers de Psychologie Cognitive
Current Psychology of Cognition

VOLUME 16 NUMBER 6 DECEMBER 1997

Target article

Commentaries

Author's response

Continuing commentary/Rationality in reasoning

Author's response