PSYCHOMETRIC *g* RELATED TO DIFFERENCES IN HEAD SIZE

Arthur R. Jensen

School of Education, University of California, Berkeley, CA 94720, U.S.A.

Summary—Scores on 17 diverse tests of cognitive abilities obtained from 82 pairs of monozygotic (MZ) and 61 pairs of dizygotic (DZ) twins were correlated with head size. A general factor, or psychometric *g*, was extracted from the tests, and *g* factor scores were found to be correlated with head size variables not only within individuals, but within twin pairs and between twin pairs. The size of the various tests' *g* loadings predicts the degree to which the tests are correlated with head size. This finding adds one more biological variable—head size and, by inference, brain size—to the list of other biological variables reported in the literature as showing a significant relation to psychometric *g*, the general factor common to all cognitive tests. Also, the varying magnitude of the mean difference between groups of white and black children on the 17 tests is related to the tests' loadings on *g*, on spatial ability, and the tests' correlations with head size.

INTRODUCTION

Since its discovery by Charles Spearman (1904), psychometric *g*, the general factor common to all tests of mental ability, has been found to be related to a number of biological variables, attesting to the view that *g* is not merely an artifact of psychometric test construction or of the methodology of factor analysis *per se*, but is a natural phenomenon that has arisen in the course of human evolution (Jensen & Sinha, 1993).

When a number of diverse tests of mental abilities are factor analyzed and a general factor, or *g*, is extracted from the tests' correlations with one another, the column vector of the tests' *g* loadings can be correlated with a parallel column vector of the same tests' correlations with some other variable. This method of correlated vectors has been applied to the vector of *g* loadings extracted from various test batteries, such as the subtests Wechsler intelligence scales, and the vector of the tests' correlations with various biological variables. Significant and substantial correlations have been found between the *g* vector and biological variables such as the heritability of individual differences in test scores (Pedersen, Plomin, Nesselroade & McClearn, 1992), the inbreeding depression of test scores (Jensen, 1983), the enhancing effect of heterosis (outbreeding) on test scores (Nagoshi & Johnson, 1986), the complexity of the average evoked potential (Eysenck, 1986), and the rate of habituation of the evoked potential (Schafer, 1985).

Also, Spearman (1927) originally suggested the idea that the variable size of the white–black mean difference on a variety of mental ability tests is related to the degree to which the various tests are loaded on the *g* factor. This idea, which I have dubbed "Spearman s hypothesis", has been investigated in many sets of appropriate data, each of which has borne out the hypothesis that *g* is the main factor in the average white–black difference on diverse psychometric tests (Jensen, 1985). But it was also discovered that, in addition to *g*, a spatial ability factor constitutes a significant part of the white–black difference on certain tests, independently of *g*.

The present study first investigates whether brain size, as inferred from head size, is also related mainly to the *g* factor of psychometric tests. Secondly, because other studies (Jensen & Johnson, in press; Jensen & Sinha, 1993; Rushton, 1992) have reported racial differences in head/brain size, the present study also asks whether head size (and by inference, brain size) is related to Spearman's hypothesis of the average white–black difference on various mental tests: Is the racial difference in test scores associated with *g* (and spatial ability) through their joint correlation with head/brain size?

A correlation between IQ and brain size, as estimated from measurements of head size, is now well established (Van Valen, 1974; Weinberg, Dietz, Penick & McAlister, 1974) and at least four studies

have shown *in vivo* correlations near 0.4 between IQ and total brain size as measured by MRI (Andreasen *et al.*, 1993; Raz, Torres, Spencer, Millman, Baertschi & Sarpel, 1993; Wickett, Vernon & Lee, 1994; Willerman, Rutledge & Bigler, 1991). Although IQ scores are highly *g* loaded, they also reflect somewhat different admixtures of cognitive abilities, such as verbal, spatial, numerical, and memory. The lead question here is whether in humans the *g* factor itself is strongly implicated in the correlation of brain size with mental ability.

Much the same question, but with respect to the correlation between behavioral capacity and brain size in rats, was asked by Anderson (1993). He found that individual differences in rats' levels of performance on four distinct 'mental' tests, when subjected to a principal factor analysis, yielded a general factor, indicating that all of the tests shared a common source of variance. In fact, the correlation matrix of the tests used by Anderson clearly meets Spearman's (1927) tetrad criterion that a single common factor accounts for individual differences in the four tests. Anderson (1993) also found that the rats' factor scores, which indicate their standing on the general factor, were more highly correlated ($r = 0.48$, $P < 0.03$) with brain weight than were the raw scores on any single test. It had been found previously that larger brains are associated with learning performance in mice, and that hybrid mice have bigger brains than their parents—another example of the enhancing effect of heterosis on brain capacity (Jensen, C., 1979).

It should be especially noted that the present study is not primarily intended as a 'twin study' in the typical sense or as a biometrical genetic analysis. The fact that the study sample consists of twins is purely incidental. This study's chief aim is simply to test the hypothesis that various cognitive tests' degree of correlation with head size (and by inference, brain size) is a function of the tests' *g* loadings. This could just as well have been determined in a sample of singletons. However, because the study sample happens to consist of twins, it is possible, as a side issue, to replicate a crucial finding of another study (Jensen & Johnson, 1994) based entirely on ordinary siblings, which shows that head size and IQ are correlated *within* families (i.e. sibships) as well as *between* families, in very large and representative samples of both the white and black populations. The fact of a significant within-families correlation between head size and IQ is not intended to prove anything about the heritability of the trait, although the *absence* of a within-families correlation definitively rules out genetic pleiotropy (i.e. variance in phenotypically different characteristic caused by one and the same gene). An intrinsic or functional connection between a physical and psychological trait is indicated only if there is a within-families correlation between them. The absence of a within-families correlation when there is a between-families correlation indicates that the between-families correlation results either from heterogeneity and common assortment of the genes that affect both of the correlated traits or from environmental factors common to all of the children in a family, or from a combination of both. But the correlation in this case does not imply any *functional* relationship between the correlated traits. In studying the causes of correlation between a physical and a mental trait, therefore, the use of within-families correlation is a powerful means of controlling the effects of many environmental factors, such as social class, styles of child-rearing, and general nutrition, that differ between families (but not within families) and may affect both of the correlated variables, causing them to be correlated in the population although they have no functional relationship. The explanation of a significant within-families correlation, therefore, is narrowed to genetic pleiotropy or to environmental factors that systematically affect both correlated characteristics in one child but not in other siblings in the same family. Of course, pleiotropy and environmental factors could both be involved. The rationale of this methodology is more fully spelled out elsewhere (Jensen, 1980), and an example of its use may be seen in a study of the correlation between myopia and IQ (Cohn, Cohn & Jensen, 1988; also see Jensen & Sinha, 1993, and Jensen & Johnson, 1994).

METHOD

Answers to the key questions of this study were sought in a set of data obtained by Osborne (1980), who gave a battery of 17 diverse cognitive tests to large samples of MZ and DZ twins and measured their body size (height, weight) and head size (length, width, circumference). Although the total sample in Osborne's study comprised 496 pairs of twins (69% white, 31% black), not all of them were given the same mental tests and physical measurements. In order to extract the best possible *g* factor

and group factors (spatial, verbal, and memory) from the largest number of tests available in any subgroup of Osborne's study, the factor analysis (and the analysis of twins) in the present study was based on only the 286 individuals (82 MZ pairs and 61 DZ pairs) for whom there were complete data on every mental test and every physical variable. The mean white–black differences on the various tests (in Table 5), however, are based on the largest N available for each test. The names of the tests and their different sample sizes are given in Table 5 and its footnotes.

Osborne (1980) described the S sample, collected in the 1970s, as follows: "Subjects for the Twin Study were drawn from public and private schools in Louisville, Kentucky, and Jefferson County, Kentucky; from public schools in Atlanta, Georgia, and the Georgia counties of Cobb, Fulton, Chatham. Walton, Madison, and Clarke; and from a small number of public schools in Indiana" (p. 34). (Several tables in Chapter V of Osborne's book give the breakdown of the total sample by age, sex, race, and twin zygosity.) The Ss ranged in age from 12 to 20 yr; their average age at the time of testing was 15.2 yr (SD = 1.5 yr).

Compared to nationwide normative data on the average IQs of whites and blacks, the present samples are not quite typical. On an IQ scale with an overall mean of 100 and standard deviation of 15, the white sample's average IQ was 104 and the black sample's was 94. The 10 points average difference is only about two-thirds as large as the white–black difference typically found on nationally standardized tests. [The mean white and black IQs in the national standardization sample of the Wechsler Intelligence Scale for Children—Revised are 102.2 and 86.4, a difference of 15.8 IQ points (Jensen & Reynolds, 1982)].

All the mental tests and physical measurement procedures are fully described elsewhere (Osborne, 1980). The names of the 17 tests, all of them well-known standardized ability tests, are listed here in Table 2. Standing height in inches was measured in stocking feet; weight was measured in street clothes without shoes. Maximum *head length*, measured with spreading calipers, is the distance in millimeters between the glabella and the farthest point on the midline on the back of the head. Maximum *head width*, measured with calipers, is the greatest transverse distance in millimeters of the head (usually found over each parietal lobe). *Head circumference*, measured with a steel tape, is the distance in millimeters from the area between the eyebrows around the maximum projection of the occiput. Twins' zygosity was determined by a combination of blood typing and other methods of diagnosing zygosity.

Prior to all data analyses, variance associated with race and sex was statistically regressed out of all the mental and physical variables as were the linear, quadratic, and cubic components of age in months.

As the vast majority of studies in which head size was correlated with mental tests did not statistically control the effect of body size in addition to age and sex, the present analysis followed suit, thereby permitting direct comparison with earlier studies (reviewed in Jensen & Sinha, 1993). Arguments against controlling for body size in studying individual differences among humans are (1) that head size and weight themselves constitute an appreciable fraction of total body height and weight, hence making the statistical removal of height and weight an over-correction, and (2) the possibility that growth in body size accommodates brain size rather than vice versa. Selective breeding of laboratory rats over 12 generations to obtain genetically 'bright' and 'dull' strains in maze learning ability steadily increased the difference between the 'bright' and 'dull' strains not only in maze learning ability but in brain weight and cranial size. Also, the two strains became increasingly different in overall body size, but only about one-third as much (in standard deviation units) as the increase in brain and cranial size (Hamilton, 1935).

RESULTS

Correlations among the physical variables are shown in Table 1.

Factor analysis

Correlations among the mental tests in the total sample were subjected to a Schmid–Leiman (1957) hierarchical factor analysis, which is generally regarded as the most appropriate method for extracting a general factor from a correlation matrix, as this method cannot possibly produce the appearance of

Table 1. Pearson correlations[a] between physical variables

Variable	L	W	C	Ht
Head length (L)	1.00			
Head width (W)	0.15	1.00		
Circumference (C)	0.66	0.46	1.00	
Body height (Ht)	0.33	0.20	0.30	1.00
Body weight	0.37	0.41	0.40	0.50

[a]Sample size = 286.

a general factor when a general factor is not in fact present in the correlation matrix (as can the first principal factor or the first principal component) (Jensen & Weng, 1994). Table 2 shows the results of the hierarchical factor analysis, in which the general factor, g, is a second-order factor, that is, the factor common to the three first-order factors, which in this battery are spatial, verbal, and memory abilities. The tests' factor loadings on the general factor are virtually identical in the white and black samples, with a congruence coefficient of + 0.99. This is typical of many other studies (see Jensen, 1985).

The g factor scores, calculated for each of the 286 individuals, were used as the dependent variable to be predicted, using both simple and multiple correlation, with the three head size measurements as the independent, or predictor, variables. The correlations were also corrected for attenuation, assuming a reliability of 0.95 for the factor scores. (In this study, however, such correction is important only for comparing the between-pairs and within-pairs correlations based on twins, as explained below. The correction for attenuation was not used in other analyses.)

Correlations of g factor scores with head measurements

Table 3 shows the simple and multiple correlations of head length (L), width (W), and circumference (C) with g factor scores. The column headed *Within Individuals* gives the correlations based on all 286 members of twin pairs treated as individuals. These correlations of L, W, and C with g factor scores are larger than such correlations reported in the literature based on IQ or various single tests, which average 0.194 (Jensen & Sinha, 1993; Johnson, 1991). In the present study, the overall average correlation of each of the 17 single tests with each of the three head size measurements (L, W, C) is 0.153.

Correlations between and within twin pairs

The theoretical importance of analyzing correlations between physical and mental traits into *between* and *within* family (or twin pair) correlations was presented in the Introduction. The between-pairs correlation (BPr) between, say, variables x and y is the correlation between the sum

Table 2. Schmid–Leiman hierarchical factor analysis: tests' factor loadings on second-order general factor (g), first-order factors (spatial, verbal, memory), and tests' communalities (h^2)

Test	g	Spatial	Verbal	Memory	h^2
PMA reasoning	0.802	0.155	0.303	− 0.021	0.763
Heim vocabulary	0.762	− 0.045	0.481	− 0.037	0.815
Cattell culture fair	0.749	0.380	0.057	− 0.043	0.707
Newcastle spatial	0.746	0.373	0.026	0.108	0.706
Arithmetic	0.669	− 0.004	0.353	0.112	0.587
PMA number	0.668	0.141	0.248	− 0.047	0.521
Surface development	0.637	0.393	− 0.030	0.006	0.568
PMA verbal meaning	0.636	− 0.056	0.383	0.120	0.544
Spelling	0.635	− 0.183	0.550	− 0.058	0.754
Paper folding	0.634	0.407	− 0.061	0.073	0.572
PMA spatial relations	0.605	0.223	0.103	0.073	0.428
Calendar	0.585	0.241	0.137	− 0.189	0.461
Form board	0.585	0.389	− 0.054	0.001	0.500
Cube comparison	0.492	0.353	− 0.044	− 0.112	0.369
Wide range vocabulary	0.448	− 0.007	0.274	− 0.069	0.287
Object aperture	0.421	0.359	− 0.135	0.082	0.352
Identical pictures	0.264	0.000	0.000	0.616	0.449
Variance (%)	38.8	7.1	6.4	2.9	55.2

Table 3. Correlation[a] of head size variables with g factor scores in individuals ($N_i = 286$), and pairs of MZ twins ($N_p = 82$), and DZ twins ($N_p = 61$)

Head Size Variables	Individuals Within	MZ twin pairs		DZ twin pairs	
		Between	Within	Between	Within
Length (L)	28** (30)	33** (33)	16 (24)	20 (21)	21 (24)
Width (W)	14* (14)	21* (21)	01 (02)	03 (04)	19 (22)
Circumference (C)	24** (25)	31* (32)	01 (01)	04 (04)	28* (32)
Multiple R					
L and W	30** (31)	38** (39)	16 (24)	20 (21)	25 (29)
L and C	29** (30)	35** (36)	21 (31)	25 (26)	28 (32)
W and C	24** (25)	32* (33)	02 (03)	04 (05)	28 (32)
L, W, and C	30** (31)	38** (39)	21 (31)	26 (26)	28 (32)

[a]Decimals omitted. Disattenuated correlations in parentheses.
*$P < 0.05$, two-tailed test. **$P < 0.01$, two-tailed test.
Significance test not applied to disattenuated correlations.

of the members of each pair on x and the sum of the members of each pair on y. The within-pairs correlation (WPr) is the correlation between the signed difference between the members of the pair on x and the signed difference between the members of the pair on y. Because WPr, being based on difference-scores, has lower reliability than BPr, comparison of BPr and WPr requires that both correlations be corrected for attenuation by the appropriate formulas (given in Jensen, 1980).

For MZ twins, the BPr of g with head size reflects genetic factors in each variable as well as any environmental factors that cause differences between families in the two traits. The WPr for MZ twins can reflect only environmental effects, pre- and postnatal, that cause one member of each pair to differ from the other in the same direction on both of the correlated variables, in this case, head size and g factor scores. For DZ twins, who have only about one-half of their genetic variance in common, the Bpr reflects both genetic and environmental differences between families in the two traits. But the WPr may reflect genetic differences between the members of each DZ twin pair as well as within-family environmental effects. Whether these environmental effects in the case of DZ twins are the same as the environmental effects that cause MZ twins to differ from each other is uncertain. [The causes of within-family nongenetic variance in IQ are considered in detail elsewhere (Jensen, in press)].

The WPr is especially important because, when twins (or full siblings) are reared together, it controls for sources of variance between families such as racial and cultural background and socioeconomic status. Also, the existence of a WPr between a physical and mental trait in the case of DZ twins (or full siblings) rules out correlation due to common assortment of genes for the correlated traits as a result of cross-assortative mating for both traits in the population. A WPr is consistent with pleiotropy, but may also have environmental causes. Note in Table 3 that the BPr for MZ twins is consistently larger than the WPr, whereas the opposite is true for DZ twins. This pattern of MZ–DZ differences suggests that part of the correlation between head size and g is attributable to the between-families (or shared) genetic component in each variable. However, the fact that the largest within-pairs multiple R between head size and g (i.e. with head length, width, and circumferences as the independent variables and g factor scores as the dependent variable) is nearly the same (0.31 vs 0.32) for MZ and DZ twins would seem to suggest that the correlation between g and head size has no genetic component and therefore could not be pleiotropic. This problematic finding is revisited in the Discussion.

In all these correlations, of course, head size is only an attenuated proxy for actual brain size. Post-mortem studies of the relation between actual brain size and external measurements of head size show an average correlation of only about 0.50 (Jensen & Sinha, 1993). Hence a rough estimate of the maximum correlation between g and brain size would be had by doubling the correlations shown in the first column of Table 3. The estimates are somewhat higher than the correlations reported between MRI measurements of brain size and IQ, but the classical IQ, with its heavy verbal saturation, probably has less g variance than the present g-factor scores. And as shown below, head size is related to the g factor and not at all to the verbal factor independent of g.

Table 4. Tests' reliability coefficients (r_{xx}), zero-order correlations and multiple correlations (R) with the three head-size variables

Test	r_{xx}	Length	Width	Circumference	R
PMA reasoning	—[a]	0.261	0.119	0.172	0.277
Heim vocabulary	0.83	0.208	0.108	0.151	0.223
Cattell culture fair	0.86	0.187	0.165	0.246	0.256
Newcastle spatial	0.94	0.246	0.162	0.210	0.277
Arithmetic	0.85	0.237	0.075	0.161	0.241
PMA number	—[a]	0.268	0.157	0.228	0.293
Surface development	0.80	0.145	0.174	0.151	0.211
PMA verbal meaning	—[a]	0.176	0.063	0.113	0.182
Spelling	0.93	0.115	0.133	0.066	0.183
Paper folding	0.73	0.276	0.167	0.238	0.304
PMA spatial relations	—[a]	0.209	0.104	0.213	0.233
Calendar	0.78	0.097	0.037	0.076	0.100
Form board	0.73	0.116	0.075	0.078	0.133
Cube comparison	0.58	0.170	0.120	0.187	0.204
Wide range vocabulary	0.88	0.173	0.091	0.194	0.204
Object aperture	0.58	0.161	0.111	0.144	0.184
Identical pictures	0.87	0.119	0.066	0.108	0.130
Mean r		0.186	0.113	0.161	0.214

[a]Not given in Osborne (1980).
Significance: For $r > 0.10$, $P < 0.05$; $r > 0.14$, $P < 0.01$; $r > 0.19$, $P < 0.001$, one-tailed test.

Correlation of tests' g loadings with head size

Table 4 shows the tests' reliability coefficients (r_{xx}), each test's zero-order Pearson correlation with each of the three head measurements, and the multiple correlation (R) of all three head-size variables with each test. Because the multiple R includes all three head-size variables, it gives the maximum correlation of the three head size measures with each of the tests. If the degree to which a given test's correlation with head size is mainly determined by the test's g loading, there should be a positive correlation between the column vector of g loadings, Vg (in Table 2), and the column vector of each of the tests' correlation with head size, VR represented by the multiple R (in Table 4). Because the g loadings are not experimentally independent variables and their population distribution is unknown, the Pearson r cannot be used to test the significance level of the relationship between Vg and VR. The proper test is Spearman's rank-order correlation coefficient (r_s), which is a non-parametric permutation test of the probability that two sets of random numbers would be as similar in rank order as the two sets of variables in question. The r_s between Vg and VR is 0.642, $P < 0.01$ (one-tailed). But we must also take into account the fact that a test's g loading and its correlations with head size are all similarly affected by the test's reliability, r_{xx}, and therefore the correlation of Vg with VR could possibly be attributed to the fact that they both reflect the same rank order as the tests' reliability coefficients. In fact, however, the rank order correlations of the of r_{xx} vector (for the 13 tests with known reliability coefficients) with Vg and with VR are small and nonsignificant ($r_s = 0.325$ and 0.121, respectively), so that partialing out the effect of reliability on the correlation between Vg and VR changes it by only $+0.001$.

This result bears out the hypothesis that the magnitudes of various cognitive tests' correlation with head size (and by implication, brain size) is positively related to the magnitudes of the tests' g loadings.

Is g the only factor that is correlated with head size, or are other factors, independent of g, also significantly correlated with head size? It turns out that none of the column vectors of the first-order factor loadings (in Table 2) is significantly correlated with the column vector of the tests' correlations with head size (column R in Table 4). For the spatial factor, $r_s = 0.267$; for the verbal factor, $r_s = 0.000$; and for the memory factor, $r_s = 0.049$. The relation of the verbal and memory factors to head size, independently of g, is clearly negligible. But the fact that the spatial vector is appreciably correlated (although nonsignificantly) with the head-size vector means that its effect on VR is to attenuate the correlation of Vg with VR. When Vs is held constant (by partialing it out of the correlation between Vg and VR), the partial r_s of Vg with VR is 0.686 ($P < 0.01$). When Vg is held constant (by partialing it out of the correlation between Vs and VR), the partial r_s of Vs with VR is 0.405 ($P < 0.05$). The multiple correlation of Vg and Vs with VR is 0.713 ($P < 0.01$). Hence head size is reflected mostly in tests that are highly loaded on both g and spatial ability.

Table 5. Mean difference between white and black groups[a] expressed in IQ units[b] and in σ units[c]

Test	W–B difference[d]	
	IQ difference[b]	σ difference[c]
PMA reasoning	14.30	1.08
Heim vocabulary	8.22	0.57
Cattell culture fair	15.10	1.16
Newcastle spatial	13.53	1.01
Arithmetic	8.67	0.60
PMA number	15.50	1.21
Surface development	10.18	0.72
PMA verbal meaning	11.50	0.83
Spelling	8.35	0.58
Paper folding	12.98	0.96
PMA spatial relations	9.70	0.68
Calendar	11.56	0.84
Form board	11.81	0.86
Cube comparison	9.70	0.68
Wide range vocabulary	7.68	0.53
Object aperture	9.57	0.67
Identical pictures	2.43	0.16
Mean difference	10.63	0.77σ

[a]Sample size: on Cattell and PMA tests, white $N = 126$, black $N = 160$. For all other tests, white $N = 540$, black $N = 237$.
[b]Raw scores on each test, with variance due to age differences statistically removed, were scaled in IQ units, with mean = 100, SD = 15 for the combined white and black samples.
[c]A σ unit is the average within-groups standard deviation.
[d]Every one of the W–B mean differences is significant ($P < 0.01$).

Head size and Spearman's hypothesis of the white–black difference

Spearman's hypothesis that the mean white–black (W–B) difference on various tests is related to the tests' g loadings is borne out in the present data by the significant rank-order correlation ($r_s = 0.484$, $P < 0.05$) between the vector of g loadings, Vg (in Table 2) on the 17 tests and the vector of mean W–B differences, $VW–B$, shown in Table 5. (The value of r_s is the same whether it is based on the IQ or the σ vector of differences, as their rank orders are identical.) When the vector of the tests' reliability coefficients is held constant (by partial correlation) the correlation between Vg and $VW–B$ is slightly increased ($r_s = 0.506$).

This correlation of 0.484 is notably smaller than the correlation generally found in 14 other large data sets, in which the average r_s is 0.62. The reason that the correlation is not larger in this (and other) studies is related to the number of tests in the particular battery that are substantially loaded on the spatial factor. Spearman's hypothesis obviously must be revised: the mean W–B difference on tests is a function both of their loadings on g and on the spatial factor. This has been evident in comparable analyses of every data set that contained a substantial spatial factor. The present battery has a larger number of spatially loaded tests than any battery previously used to examine Spearman's hypothesis.

Hence the vector of loadings on the spatial factor, Vs (in Table 2), is significantly correlated with the vector of W–B differences ($r_s = 0.532, P < 0.05$). Partialing out the vector of reliability coefficients slightly increases the correlation ($r_s = 0.595$).

If both the g and spatial (s) factors are independently related to the W–B difference, what is their joint relation to the W–B vector? This was determined by combining the factor loadings of g and s on each test and obtaining the rank-order correlation between the $g + s$ vector and the W–B vector. This r_s is 0.744 ($P < 0.01$); partialing out the tests' reliability coefficients increases this r_s to 0.808.

Finally, if head (or brain) size is among the causal variables that mediate the relation between the $g + s$ vector and the vector of the mean W–B differences on various tests, one should expect to find a correlation between the vector of the tests' correlations with head size, VR (in Table 4) and the vector of W–B differences on the tests (Table 5). This correlation turns out to be significant ($r_s = 0.533$, $P < 0.05$); with the vector of the tests' reliability coefficients partialed out, the $r_s = 0.715$ ($P < 0.01$).

The vectors of the verbal and memory factors (Table 2) are negatively, but not significantly, correlated with the vector of W–B differences on the various tests. For the verbal factor, $r_s = -202$; for memory factor, $r_s = -0.116$.

It is important to note that none of the factor loadings in Table 2 and none of the correlations in Tables 1, 3, and 4 [and hence none of the vectors (except the vector of mean W–B differences in Table 5) that enter into these analyses] can possibly reflect the effect of race differences on any of the variables, because the dichotomous variable of race (quantitized as W = 1, B = 0) was regressed out of all the test variables and physical variables of every S prior to performing the above analyses. This is equivalent to saying that all of the correlations and factor loadings are the average *within* racial-groups statistics. (Similarly, the effects of age and sex have been removed from all the variables.) Race differences appear only in Table 5.

A multiple rank-order correlation (R_s) that includes the $g + s$ vector along with the vector of head-size correlations as predictors of the W–B vector turns out to be 0.747 ($P < 0.01$); with the vector of reliability coefficients partialed out, this correlation becomes 0.828 (after correction for shrinkage $R_s = 0.800$). The fact that this shrunken multiple R_s (0.800) scarcely differs from the simple r[ins]s (0.807) of the $g + s$ vector with the W–B vector implies that the head-size vector adds virtually nothing to the prediction of the W–B differences in the various tests over and above what is predicted by $g + s$. This is consistent with the results of a study which found that groups of white and black children who were matched on IQ (also on age, height, and weight) did not differ in head size, although the total white and black samples, which differed about 1σ in IQ, differed significantly ($P < 0.001$) in head size (W > B), with age, height, and weight statistically controlled (Jensen & Johnson, 1994).

DISCUSSION AND CONCLUSIONS

An analysis of MZ and DZ twins extends a previous finding of a significant correlation between g factor scores and head size both within individuals and within pairs of dizygotic twins (Johnson, 1991, p. 327). Theoretically a within-pairs correlation implies, but does not necessarily prove, a pleiotropic relation between two highly heritable traits, such as g and head size presumably are. Environmental factors, such as nutrition, could also affect both variables. But the pattern of between- and within-pairs MZ and DZ twin correlations reported in the present study (Table 3) is puzzling in this regard. The within-pairs correlation between head size and g for MZ twins can be only a purely nongenetic correlation. The within-pairs correlation for DZ twins theoretically could be a pleiotropic genetic correlation, but this is unsupported by the fact that the obtained correlation in the DZ twins is no larger than the purely nongenetic correlation in the MZ twins. Could it be the case that the correlation between head size and g is wholly nongenetic, despite the substantial heritability of both g and head size? This would seem most surprising if true. But for lack of adequate statistical power, it can be neither proved nor disproved by the present data. Ideally, the data of the headsize $\times g$ covariance would be subjected to a biometrical genetic analysis, statistically testing the goodness-of-fit to alternative models of the relations between the shared and nonshared genetic and nongenetic components of the total covariance. Unfortunately, the present sample sizes of MZ and DZ twins are insufficient to statistically discriminate with much confidence between plausible models having this many parameters. It would involve, in effect, the partitioning of a phenotypic correlation (or covariance) which, given the present degrees of freedom, is itself barely significant, and then testing the significance of the differences between the partitioned components. The null hypothesis would inevitably be overworked. The sample sizes required for sufficient statistical power to detect various components of genetic and environmental variance in a biometrical analysis is discouragingly much greater than is often supposed (see Neale & Cardon, 1992, Chap. 9). A biometrical model-fitting analysis of the genetic and environmental components of just the head size measurements (combined by formula to estimate cranial capacity) in Osborne's (1980) study, based on a larger sample than was possible in the present study, found heritabilities of cranial capacity ranging between about 0.40 and 0.50, with blacks showing somewhat lower heritability than whites (Rushton & Osborne, in press). But a purely environmental model could not account for the phenotypic variance in cranial capacity of either blacks or whites; genetic factors are clearly required. Osborne's data also show mental test score heritabilities similar to those generally found in twin studies and not differing significantly between whites and blacks.

It has been observed that head size (and by implication, brain size) is correlated to varying degrees with different psychometric tests of mental ability. It is also well established that various tests of mental

ability have different factor structures, that is, the independent sources of variance that all tests, or certain types of tests, have in common, but to varying degrees. The factor structure of various mental tests typically consists of g, or the general factor common to all cognitive tasks, one or more group factors (e.g. verbal, spatial, number, etc.), and factors that are specific to each test, in addition to error variance due to random errors of measurement. The main question addressed by this study, therefore, concerns which factors in various tests are the most strongly related to variation in head size. The answer would suggest which psychometric factors most closely reflect the biological substrate of human mental abilities, whether that substrate is more affected by genes or by environment.

It was found that the degree to which various tests are correlated with head size is primarily a function of the tests g loadings, that is, the larger a tests' g loading, the greater is its correlation with head size. Also, tests that include a spatial factor are somewhat more highly correlated with head size than would be predicted from the tests' g loadings alone. The verbal and memory factors, independently of g, were negligibly related to head size. The reason that many studies have shown significant correlations of IQ with head size is because IQ tests are very highly g loaded, whatever other factors they may contain.

A related question addressed by this study concerns Spearman's hypothesis that the variable size of the mean white–black (W–B) difference on diverse mental tests is a function of the tests' g loadings, and whether this relationship is mediated to some degree by whatever brain mechanisms are associated with brain size, for which head size is merely an attenuated proxy. It was found that tests' loadings on both g and the spatial factor predicted the size of the W–B difference on the tests. Also, the degree to which the various tests are correlated (*within* racial groups) with head size was found to predict the sizes of the W–B difference on the tests. Combining the vectors of tests' g loadings, spatial factor loadings, and head size correlations in a multiple regression to predict the size of the W–B differences on the diverse tests yielded a multiple correlation (after corrections for measurement error and shrinkage) of 0.80. This finding suggests that some aspect of the biological substrate associated with head/brain size is involved in the mean W–B difference in mental test scores, particularly on tests with large loadings on the g and the spatial factors. The present analysis, however, does not resolve the question of the relative magnitudes of possible genetic and environmental effects on the biological substrate. Other kinds of analysis based on quantitative genetic models would be required to tackle this question, but much larger samples of MZ and DZ twins would be needed than were available for the present study.

Probably future investigations of the hypothesized relation of tests' g loadings (and other psychometric factors) to their correlations with brain measurements will be made more directly by means of magnetic resonance imaging, focussing not only on overall brain size but also on specific brain structures.

REFERENCES

Anderson, B. (1993). Evidence from the rat for a general factor that underlies cognitive performance and that relates to brain size: intelligence? *Neuroscience Letters, 153*, 98–102.

Andreasen, N. C., Flaum, M., Swayze II, V., O'Leary, D. S., Alliger, R., Cohen, G., Ehrhardt, J. & Yuh, W. T. C. (1993). Intelligence and brain structure in normal individuals, *American Journal of Psychiatry, 150*, 130–134.

Cohn, S. J., Cohn, C. M. G. & Jensen, A. R. (1988). Myopia and intelligence. *Human Genetics, 80*, 53–58.

Eysenck, H. J. (1986). The theory of intelligence and the psychophysiology of cognition. In Sternberg, R. J. (Ed.), *Advances in the psychology of human intelligence* (Vol. 3). Hillsdale, NJ: Erlbaum.

Hamilton, J. A. (1935). The association between brain size and maze ability in the white rat. Unpublished doctoral dissertation, University of California, Berkeley.

Jensen, C., (1979). Learning performance in mice selected for brain weight. In Hahn, M. E., Jensen, C. & Dudek, B. C. (Eds.). *Development and evolution of brain size*. San Diego, CA: Academic Press.

Jensen, A. R. (1980). Uses of sibling data in educational and psychological research. *American Educational Research Journal, 17*, 153–170.

Jensen, A. R. (1983). The effects of inbreeding on mental ability factors. *Personality and Individual Differences, 4*, 71–87.

Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests: Spearman s hypothesis. *Behavioural Brain Science, 8*, 193–219.

Jensen, A. R. (in press). The puzzle of nongenetic variance. In Sternberg, R. J. & Grigorenko, E. L. (Eds.), *Intelligence. heredity and environment*. Cambridge, MA: Cambridge University Press.

Jensen, A. R. & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence, 18*, 309–333.

Jensen, A. R. & Reynolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. *Personality and Individual Differences, 3*, 423–438.

Jensen, A. R., & Sinha, S. N. (1993). Physical correlates of human intelligence. In Vernon, P. A. (Ed.), *Biological approaches to the study of human intelligence*. Norwood, NJ: Ablex.

Jensen, A. R. & Weng, L-J. (in press). What is a good *g*? *Intelligence, 18*, 231–258.

Johnson, F. W. (1991). Biological factors and psychometric intelligence: a review. *Genetic Social and General Psychological Monographs, 117*, 313–357.

Nagoshi, C. T. & Johnson, R. C. (1986). The ubiquity of *g*. *Personality and Individual Differences, 7*, 201–207.

Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Boston, MA: Kluwer Academic.

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

Pedersen, N. L., Plomin, R., Nesselroade, J. R. & McClearn, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science, 3*, 346–353.

Raz, N., Torres, I. J., Spencer, W. D., Millman, D., Baertschi, J. C. & Sarpel, G. (1993). Neuroanatomical correlates of age-sensitive and age-invariant cognitive abilities: An *in vivo* MRI investigation. *Intelligence, 17*, 407–422.

Rushton, J. P. (1992). Cranial capacity related to sex, rank, and race in a stratified sample of 6,325 U.S. military personnel. *Intelligence, 16*, 401–413.

Rushton, J. P. & Osborne, R. T. (In press). Genetic and environmental contributions to cranial capacity in black and white adolescents. *Intelligence*.

Schafer, E. W. P. (1985). Neural adaptability: A biological determinant of *g* factor intelligence. *Behavioural Brain Science, 8*, 240–241.

Schmid, J. & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53–61.

Spearman, C. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Spearman, C. (1927). *The abilities of man*. London: Macmillan.

Van Valen, L. (1974). Brain size and intelligence in man. *American Journal of Physical Anthropology, 40*, 417–424.

Weinberg, W. A., Dietz, S. G., Penick, E. C. & McAlister, W. H. (1974). Intelligence, reading achievement, physical size, and social class. *Journal of Pediatrics, 85*, 482–489.

Wickett, J., Vernon, P. A. & Lee, D. H. (1994). *In vivo* brain size, head perimeter, and intelligence in a sample of healthy adult females. *Personality and Individual Differences, 16*, 831–838.

Willerman, L., Rutledge, J. N. & Bigler, E. D. (1991). *In vivo* brain size and intelligence. *Intelligence, 15*, 223–228.