Bilag A (4)

Sex-related differences in general intelligence *g*, brain size, and social status

Helmuth Nyborg *
International Research Center
for Psychoneuroendocrinology
Department of Psychology
University of Aarhus
Risskov, Denmark

Revised version 050104

* Corresponding author. Tel.: +45 87680456.
E-mail address helmuth@psy.au.dk (H. Nyborg)

Abstract

The question of a sex difference in intelligence and its relationship to social status has long divided the experts. IQ researchers sum standardized subtest scores to calculate intelligence in general, and find that males outscore females by about 3.8 points, whereas factor analysts derive the *g* factor scores from intertest-correlations and find no consistent sex differences in general intelligence. The latter finding is puzzling. Males have larger average brains than females, and brain size correlates .30 - .40 with *g* (and IQ). From this evidence, males "ought" to have higher *g* score than females.

The present study addressed this paradox by testing four hypotheses: (1) Inadequate definitions and analyses explain why researchers get inconsistent results, (2) The proper method identifies a male *g* lead, (3) The larger male brain "explains" the male *g* lead, 4) Classic Distribution Theory illustrates how the average male SD *g* score advantage and the wider male SD distribution score transform into an exponentially increased male-female ratio at the very high end of the *g* distribution, which largely explains male dominance in society.

All four hypotheses obtained support. This outcome provides part of the explanation why a few percentiles of the male population come to dominate the upper educational, occupational, and political strata in all known societies, and how relatively few males come to occupy 88% of the power positions in contemporary equality-oriented Denmark. The support for hypothesis 3 suggests that the link between brain size, general intelligence, and social dominance is partly biological.

1. Introduction

Experts have long discussed the existence, nature, size, and importance of a sex difference in overall intelligence. Some researchers (e.g. Lynn, 1994; 1997; 1999; Lynn, Irwing & Cammock, 2002) find that males outscore females by about 3.8 IQ points, whereas the majority finds no sex difference (e.g. Brody, 1992; Halpern & LaMay, 2000; Jensen, 1998). On the other hand, few experts dispute the existence of marked sex differences in group factors like verbal and spatial abilities, where females excel in the first and males in the latter, in particular when verbal fluency and 3-D abilities are involved.

The disagreement is very confusing for two theoretical reasons. First, males are grossly over-represented at the higher ranks of education, research, occupation, and political power structures, categories that all call for high capacity to deal with complexity, which is just another way of defining general intelligence. Second, males have, on average, larger brains than females, and brain size correlates positively with intelligence. Yet, the empirical evidence for the male advantage in overall intelligence is equivocal.

The purpose of the present work is to address these paradoxes by testing four hypotheses: (1) Ambiguous definitions of intelligence and inadequate use of analytic methods explain the current disagreement among the experts, (2) The proper analytic approach will identify a male lead in general intelligence *g*, (3) The larger average male brain explains the average male *g* lead, and, finally, (4) Classical Distribution Theory illustrates how a small male mean *g* SD score advantage and a wider male SD dispersion score result in an exponentially increased male-female ratio at the very high end of the *g* distribution. The increasingly unequal ratio of more high *g* males than females at the upper end of the population distribution may explain why males (always, according to Goldberg, 1977) dominate the intellectually most demanding top occupational and political strata.

2. Hypothesis 1: Ambiguous definitions of intelligence and inadequate use of analytic methods explain the empirical inconsistency

2.1 The IQ position

Some researchers (e.g. Lynn, 1994; 1997; 1999) argue on theoretical ground that males ought to score a higher average IQ than females, and their argument can be expressed in the form of two syllogisms (Nyborg, 2002). First, brain volume correlates with IQ. Males have on average larger brains than females. Ergo: Males have a higher mean IQ than females. Second, job status and income correlate with IQ. Males have on average higher job status and income than females. Ergo: Males have a higher mean IQ than females. Several empirical studies support their arguments. By averaging a number of studies, Lynn thus found a mean male lead of 3.8 IQ points.

There is a problem with this finding, however. The total summed IQ score may be affected to an unknown degree by test bias. Female IQ will, for example, be favoured by an overweight of subtests tapping verbal abilities, and male IQ will be favoured by a spatial ability subtest bias. This means that a sex difference in IQ – a synonym for intelligence in general - may either reflect a test bias or a real sex difference, but we would not know the difference by just summing standardized subtest scores (Jensen, 1998).

2.2 The g position

A methodologically more accurate approach is to factor analyse the intertest-correlations among subtests and derive *g* factor scores. The *g* factor scores reflect general intelligence and show higher reliability and validity than IQ scores. Several types of factor analyses are commonly in use, e.g. principal factor (PF), principal component (PC), or hierarchical factor analysis (HFA). The only type not suited for *g* analysis is Thurstone's simple structure approach. This is because it mathematically prevents the *g* factor dimension from appearing by dispersing its variance across the other dimensions. A detailed discussion of various ways to extract *g* is given by Jensen (1998, chapter 4: Models and characteristics of *g*).

However, factor analysis provides inconsistent results with respect to sex differences: Females outscore males in some studies, males do better in other studies, and the remaining studies show no sex difference in *g* (Aluja-Fabregat et al., 2000; Colom et al., 2002; Colom & García-López, 2002; Colom et al., 2000; Jensen, 1998).

There is a methodological reason for this inconsistency. The common PF and PC analyses may result in a *g* score that is contaminated to a small extent by the sex differences in the group factor dimensions or by test specificity (Nyborg, 2001; 2002; 2003). This is not a problem in studies of large group differences in *g*, such as among races or social classes, where contamination would, at most, lead to an underestimation. However, if the task is to single out a perhaps small sex difference, even a slight contamination of *g* by group factors or test specificity may either drown a real difference and lead to a Type II error, or may erroneously indicate a male or a female superiority and lead to a Type I error. Jensen (1998) demonstrated the danger by factor analysing massive data from five separate studies using inclusive test batteries and representative samples. He found a female PC1 *g* lead in one test, a male PC1 *g* lead in another test battery, and no sex differences in the remaining studies (ibid. pp. 538-543. The inescapable conclusions were that sex differences in *g* are either totally nonexistent or of uncertain directions and of inconsequential magnitude, and further that the differences observed must be attributable to factors other than *g*.

However, when Jensen eliminated the unusually large number of test items favouring females in the General Aptitude Test Battery, and repeated the factor analysis, the female *g* lead disappeared! This suggests that the female *g* superiority was an artefact of a test bias that favoured females and was reflected in the PC1 *g* score. Obviously, by the same token, a male test bias would erroneously have produced male *g* superiority. The exercise raises doubt about the reliability of *g* studies of sex differences, and makes the practice of averaging variously contaminated *g*s from assorted studies a non sequitur.

2.3 Conclusion

Hypothesis 1 – that the disagreement among experts about the existence of a sex difference emanates from conceptual and methodological difficulties – thus found support.

3. Hypothesis 2: The proper analytic approach will identify a male lead in general intelligence, *g*

3.1 Introduction

Jensen's (1998) review of studies on sex differences indicated that most IQ and *g* studies suffer from non-representative sampling, and that the extensive inclusion of above average students increases the likelihood of finding a male superiority because of the wider male dispersion of IQ or *g* scores. Further problems are that many studies incorporate less than the minimum of 9 tests needed for proper factor analysis, and test batteries often show too little internal diversity to tap into the wide range of abilities needed to permit extraction of at least three different group factors.

These methodological problems lead Nyborg (2003, pp. 199-201) to develop a rough and ready grading scale for ranking current studies of sex differences in *g* in terms of quality. Briefly, a study earns one point for unbiased sampling, one for incorporating nine or more tests in the battery, one for securing sufficient test diversity, one for applying HFA analysis and performing an orthogonal Schmid-Leiman transformation (Schmid & Leiman, 1957), one for inserting the point-biserial correlations (see later) into the inter-test correlation matrix for co-factoring and, finally, one point for testing whether sex loads statistically significant on *g*. No points are given for the inclusion of correlated vector analysis in sex difference studies, for reasons given later. Studies of sex differences in *g* earning < 5 points are deemed untrustworthy, as the risk of committing Type I or Type II errors is too large.

There are only two current studies that earn the maximum 6 points on the quality scale: The present study, and one by Colom et al., (2002). Data from these two studies are analysed here, in that order, to critically test hypothesis 2, which says, that the proper analytic approach will identify a male lead in general intelligence *g*.

3.2 The present study

3.2.1 Method.

An exceptionally careful sample selection began in the late 1970es with a computer search in the Danish Folkeregister for every twentieth child that was either 8, 10, 12, 14 or 16 years old (± 0.5 year), was either a boy or a girl, and was attending a school in the Skanderborg communal district, situated either at the countryside, in a suburb, or in a larger city. Information about the socio-economic status of the parents, defined by father's occupational status, was also collected and categorized at five levels. If the twentieth child, or its parents, refused to participate in the 20+ years cohort-sequential study, the twenty-first (or in two cases the twenty-third) child on the computer list was invited. No particular pattern of reasons for refusing to participate could be spotted in retrospect. Five age group categories, 8, 10, 12, 14, or 16+ year, were established on basis of this preliminary search protocol. The age group distributions of sex, geographical characteristics, and socio-economic background of the children were inspected when about 50 percent of the children were tested, and the age categories were then supplemented to full capacity, so that each age category finally included a total of 15 boys and 15 girls. Great care was taken to ensure that all categories ended up being representative with respect to the general Danish socio-economic population distribution while also conforming to the nationwide proportional representation of rural, suburban or city residency and school attendance. The total cohort-sequential study included two cross-sectional phases – 1976 (including only some cognitive testing) and 1996- 2000+ (full study), and a full longitudinal study spanning 1981-2000+, which included two repeatedly tested groups of 30 boys and girls, each, and a control group of 30 boys and girls that were tested at age 10 and 16+ years, only. The present analysis is based on the sub-sample of an equal number of all males and females that had been tested only once and for whom

WAIS and all the other test data were available (see below). This sub-sample consisted of 26 females (mean age 17.34, SD 1.87) and 26 males (17.45, SD 1.78 SD).3.2.2. Test battery

The 52 subjects went through a very large battery of 20 ability tests, differing widely in content area. Negative scores were inverted. Briefly, the tests are 1) The Rod-and-Frame test (RFT) Frame Dependence (signed errors, inverted: Nyborg & Isaksen, 1974), 2) RFT Response Variability (inverted: ibid.), 3) RFT Field Dependence (unsigned errors, inverted: Asch & Witkin, 1948), 4) The Embedded-Figures Test (seconds/figure, inverted: Witkin, 1950), 5) The Money Left-Right Discrimination Test (errors, inverted: Money, 1965), 6) Mental Rotation Test (nos. of figures found, corrected for guessing: Vandenberg & Kuse, 1978), 7) Tapping Test, Left Hand (max. nos. of taps during 2x30 seconds), 8) Tapping Test, Right hand (max. nos. of taps during 2x30 seconds), 9) Oral Fluency (name as many animal names as possible, beginning with F within a minute), 10-20) All 11 Wechsler Adult Intelligence Scale raw scores (WAIS: Wechsler, 1958), that is, 10) WAIS Information, 11) WAIS Comprehension, 12) WAIS Arithmetic, 13) WAIS Similarities, 14) WAIS Digit Span, 15) WAIS Vocabulary, 17) WAIS Digit Symbol, 18) WAIS Picture Completion, 19) WAIS Block Design, 20) WAIS Picture Arrangement, and 21) WAIS Object Assembly.

3.2.2 Analysis

A preliminary PC analysis of the separate male and female data was first performed to check for identical factor structure in male and female data. The congruence coefficient was fairly close to unity (.947), which indicates virtual identity in the *g* factor structure for males and females.

According to Jensen (1998, p. 538), "The best method for determining the sex difference in psychometric *g* ..." is to first fit the d effects (the sex difference on each subtest, weighed by the common male-female variances) into the formula (ibid. p. 542, note 9), to calculate the point-biserial correlations (r_{pbs}, indicating the extent to which sex, as a dichotomous variable, loads on the metric sex differences), and then to insert the twenty r_{pbs} in the subtest inter-correlation matrix, and factor analyse them together with a very large number of highly varied tests. It has been confirmed experimentally that the inclusion of the sex r_{pbs} in the correlation matrix has no effect on the factor structure and only negligible effects on the subtest's g loadings (Jensen, 1998, p. 542, note 9).

The present study used an HFA analysis with the Schmid-Leiman (SL) transformation. The HFA/SL analysis provides estimates of test specificity at the lowest level, of group factors at the next level, and of general intelligence *g* at the highest second or third order level. This higher-order *g* factor show high reliability and heritability, little dimensional contamination, have many close biological and brain correlates, and show better predictive validity than other estimates of overall intelligence (Jensen, 1998). The loading of sex on *g* was tested for one-sided significance at the 2.5% level.

The HFA/SL analysis permitted extraction of one second-order *g* factor. The analysis also turned out seven first-order group factors, reflecting common sex differences. Their developmental course will be discussed elsewhere in connection with cohort-sequential data for the 325 8-14 year old boys and girls. Suffice it here to mention that no significant sex difference in *g* was seen in 8-14 years children, but there were changes in the appearance or disappearance of certain group factors.

As a matter of principles, the results of a correlated vector analysis (Jensen, 1998, appendix B) are also presented here, even if this type of analysis is a priori deemed inadequate for the study of sex differences (Nyborg, 2003, chapter 10). The purpose is basically to check whether the sex difference in *g* is a "Jensen Effect", that is, if the sizes of d differences in the vector of the 20 tests correlate with the sizes in the vector of the respective tests' *g* loadings, even if this method works hard against finding statistical significant relationships (see below).

3.2.3. Results

Table 1 first outlines the type of tests used. Column two gives the observed sex difference on each test (d effects, where a minus sign indicates female superiority) and the average effect size and its average IQ equivalent. Third column gives the r_{pbs} (adjusted for unequal SDs according to the formula given by Jensen, 1998, p. 542, note 9). The fourth column outlines the result of the hierarchical factor analysis, that is, the various tests' *g* loadings on the second order *g* factor dimension. Last row gives the average *g* load. As predicted, co-factored sex loads positively (r_{pbs} = .272) on the *g* factor. The coefficient is statistically significant ($p = .026$, one-sided), even in this small sample.

Insert table 1 and 2 about here

Table 2 shows the results of the correlated vector analysis. The *g* factor loadings of the 20 tests were corrected for attenuation by dividing each subtest's *g* loading by the square root of that subtest's reliability coefficient, as were the d values. The reliability coefficients for the first ten tests have not been determined directly in a comparable subject sample. Therefore, each test's communality, i.e. the proportion of its total variance accounted for by the common factors, was used as a lower-bound estimate of the test's reliability. This procedure actually works against the hypothesis of the sex load on *g*, because it removes some part of the *g* vector correlation with the d vector, but it is preferable to no control at all (Jensen, 1998). The reliability coefficients for the ten WAIS tests were taken from Jensen (1980).

The correlation of the *g* vector with the d vector amounts to Pearson $r = .455$, $p = .022$, one-sided. However, outliers or other peculiarities in the *g* factor loading or d scales could produce a bias that would remain unrecognised unless also a Spearman's rank-order correlation (r_S) is calculated, and its size compared to that of the Pearson r. The $r_S = .268$ ($p = .127$) is about half the size of r. This discrepancy suggest a bias, and the Spearman rank-order r_S is accordingly used here as the standard. In other words, the observed sex difference in *g* is a weak Jensen Effect not reaching statistical significance. However, it is worth noting that the degrees of freedom are restricted to the number of tests, not to the number of subjects, and this makes the correlated vector analysis highly susceptible to a Type II error, that it, to rejecting a real sex load on *g*.

3.3 The Colom et al. (2002) study

3.3.1. Introduction

There is only one other study in the entire literature on general intelligence, taking a similar high quality methodological approach (Colom et al., 2002). Paradoxically, this study identified a statistically highly significant sex difference in general intelligence *g*, but concluded that there is: "Null sex difference in general intelligence"!

3.3.2. Method

The study included a large sample of 703 females and 666 males, participating in the Spanish standardization of the WAIS-III test, and found a male IQ advantage of 3.6 points in intelligence in general – close to the average 3.8 IQ points found by Lynn (1994; 1997) and to the 3.94 equivalent IQ points in the present study (Table 1). More importantly, Colom and colleagues fitted the r_{pbs} among each subtests' score and the dichotomous sex variable into the matrix of sub-test inter-correlations, performed a HFA/SL type factor analysis of the full matrix, and found that sex loaded .159 on *g*. Unfortunately, this high-quality sex-loaded *g* value was combined with several low-quality sex-loaded *g* values, and the widely differing values cancelled out each other, so the resulting average load of sex on *g* approached zero, and the researchers could only conclude that there is "Null sex difference in general intelligence".

3.4. Re-analysis of the Colom et al. study (2002)

A methodologically more correct approach is to test the statistical significance of the observed sex load on *g* in the Colom et al. (2002) study. This shows that the observed male lead in *g* is highly significant (N = 1,369; p < .0001, Fisher z = .16036 (Nyborg, 2003).

This means that a full-size representative standardization study provides a largely uncontaminated HFA/SL *g*, and confirms the existence of a significant sex difference in general intelligence *g*, even if it was based on data from the WAIS-III test battery that was originally derived of test items with a large sex bias during its construction.

3.5 Conclusion

The only two existing high quality studies verify independently, that there is a moderate average *g* difference in male favour, thus confirming hypothesis 2.

4. Hypothesis 3: The larger male brain explains the male lead in *g*

4.1 Introduction

Many studies (Anderson, 2003; Ankney, 1992; 1995; Gignac, Vernon, & Wickett, 2003; Haier, 2003; Lynn, 1994; 1999; Rushton, 1992; and Rushton & Ankney, 1996) demonstrate that males have, on average, a larger brain (d = .30 to .35) that contains about 15% more neurons than the female brain (Packenberg & Gundersen, 1997). Moreover, brain size correlates.10 - .45 with IQ. The marked differences in correlations reflect measurement error plus the type of measure used – from a rough approximation of brain size by simply taping head circumference to precisely scanned brain volume – but the rule seems to be: The more exact the measure, the higher the correlation.

Lynn actually used the sex difference in brain size to predict the male IQ lead, and failed only by a mere -.2 IQ point (see table 3, from Nyborg, 2002).

Insert table 3 about here

Given a male lead in brain size of d = .78, and a mean correlation of .35 between brain image size and IQ (Rushton & Ankney, 1996, though Gignac, Vernon, & Wickett, 2003, suggest a mean size – IQ correlation of .40, and which Schoenemann et al., 2000, raises further to .45 for the brain volume – *g* relationship), the multiplication of the male lead in brain size with the brain size - IQ correlation gives an SD of .27, which, when multiplied by 15 translates into a male lead of 4.05 IQ points. In other words, Lynn's theoretical prediction of intelligence from brain size matches the empirically observed male average IQ lead of 3.8 quite well, but the major problem with this calculation is, that it is based on IQ scores or intelligence in general, rather than on the more precise and less contaminated HFA/SL general intelligence *g* measure.

The present study allows us to test whether Lynn's prediction based on intelligence in general can be generalised in terms of general intelligence *g*, and it allows for two different methodological approaches, both exemplified in table 3. The first is to multiply the observed sex difference in head circumference (d = .87 in this study) with the correlation coefficient between head circumference and the g factor scores (.34), which gives a predicted male lead in *g* of .30 (or 4.44 equivalent IQ points). The second approach takes point of departure in the already sex loaded point-biserial correlation of r_{pbs} (.272, see table 1), and derive its corresponding d value by the formula given by Jensen (1998, p. 543, note 12). This gives a sex difference in male favour of d = .565 (see E, table 2). Multiplying d by 15 gives an IQ lead of 8.48 equivalent points – almost double the value of the male 4.48 IQ point lead predicted on basis of the *g* factor scores.

However, the decisive reason for using the r_{pbs}, instead of the *g* factor scores as the best point of departure, is that the *g* factor score of a person will necessarily be contaminated to some extent by group factors and test specificity, as it is based on a *g*-weighted mean of that individual's standardised scores on each of the subtests, which may increase or decrease the mean sex difference as a function of the type of subtests included.

With respect to appraising how realistic is the large male-female difference in *g* of .272 found in the present study, it is sobering to remember that sex loaded considerably less (i.e. .159) on *g* in the Colom et al., (2002) study. This corresponds, according to the formula by Jensen (1998, p. 543, note 12) to a sex difference in male favour of $\underline{d} = .322$, or a male *g* lead comparable to 4.83 IQ points. This sex difference may be more realistic than the 8.48 IQ points found in the present study, as it was obtained in a large standardisation study. On the other hand, Colom and colleagues used the Spanish version of the Wechsler Adult Intelligence Scale III, and Wechsler actively wrote off test items with large sex differences during the construction of his test, so that the final selection of sub-test missed the most difficult 3-D visuo-spatial visualisation and orientation tests absent favouring males. In contrast, the present test battery represents 20 highly varied categories, giving rise to no less than 7 group factors. Obviously, further studies are nevertheless needed to establish the final size of the sex difference in *g*.

4.2. Conclusions

The analysis of brain size - intelligence relationships produced three sets of data: 1) Lynn (1999) over-predicts the male IQ lead from brain size by .2 points, and underestimates the actual average sex difference by 1.7 IQ points (5.5 - 3.85), based on biased data; 2) The present study under-predicts the male lead in *g* from head size by .07 SD (.37 minus .30), or 1.11 equivalent IQ points when using slightly contaminated *g* factor scores, but finds a sex loaded $r_{pbs} = .272$, reflecting a sex difference of $\underline{d} = .57$, or 8.48 IQ equivalent points. Finally, 3) Colom et al., (2002) find a sex load of .159, or 4.83 equivalent IQ points,.

These results provide concurrent support for the third hypothesis, namely, that the larger average male brain explains a significant part of the average male lead in *g*.

5. Hypothesis 4: Classical distribution theory illustrates how the average male *g* advantage and the wider distribution transform into an exponentially increased male-female ratio at the very high end of the *g* distribution, which might explain male dominance in society

5.1. Introduction

A male lead in *g* makes it plausible, at least in part, why the upper few percentiles of a population with the highest *g* consist almost entirely of males that eventually come to dominate the upper educational, occupational, and political strata, because there is very strong selection for high *g* in such areas (Gottfredsson, 2003), which, I would add, most likely works largely irrespective of sex.

5.2. Classical Distribution Theory

A plot of the separate distributions of male and female *g* score illustrates the effects a moderate average male lead in *g* and broader distribution of scores will have on the male-female ratio at various points on the presumed normal distribution of *g* scores. The 26 males in the present study obtained an average *g* score of .36 SD (dispersion SD = 1.06), and the 26 females scored g = -.01 SD (dispersion SD = .74).

This gives a sex difference in *g* score of .37 SD (table 2) which, according to Classical Distribution Theory, results in a modest increase in the male-female ratio at the high (and low) end of the *g* distribution. However, the difference in male-female dispersion SD scores has a more pronounced effect on the male-female ratio at the high and low ends of the *g* scale. The combined effects of the mean and dispersion differences are illustrated graphically in figure 1.

Insert figure 1 about here

The present data indicate that the theoretical ratio of males to females with *g* = 1 SD (equal to IQ 115) is 1.48:1. For *g* = 2 SD (IQ 130) the male-female ratio grows to 8.43:1. For *g* = 3 SD (IQ 130) the male/female ratio is a staggering 122.85:1. In other words, the male/female ratio increases exponentially as a combined function of the modest male lead in *g* and the larger male dispersion in *g* scores, but the sex difference in dispersion SD has a much larger effect on the sex ratio than a difference in score SD.

5.3. Survival of the welfare state

The proportions of people with a *g* factor score above 2 or 3 SDs amounts to a 2.15% and .13%, respectively, which, in terms of Denmark with a population of 5.25 million people, we talk about 11.288 or 6.825 individuals, respectively. This suffices to meet the high intellectual demands for effectively running a small modern well-fare state, when combined by the supportive assistance of a much larger number of moderately intelligent people below to sustain the necessary supportive infrastructures, and under the provision of a not too large proportion of very low *g* people, unable to cater for themselves and their offspring.

The Very High-end Male *g* hypothesis (Nyborg, 2003, p. 215) posits that the overweight of extremely high *g* males helps explain why males have been able to dominate almost all the higher spheres of the socio-economic ladder. A recent analysis of the power structure in long-time-hard-equality-seeking Denmark finds that no less than 88 percent of the most influential people in society are males (Christiansen, Møller, & Togeby, 2001). Apparently, the offensive Danish state and union supported politically, educationally, and occupationally mediated campaigns, running since the 1960s, have failed to change the picture.

Obviously, personality factors like empathy, aggression, dominance, and persistence obviously are also part and parcel of the story, in addition to the fact that many career women will be penalized for childbearing. Then again, recent studies confirm that *g* is the best single predictor of occupational status and income (e.g. Nyborg & Jensen, 2001), and of how people perform in life at large (e.g. Gottfredson, 2003).

5.4. Conclusion

The combined evidence supports hypothesis 4, that is, the heavy overrepresentation of very high *g* males largely explains why males typically out-compete females at the highest steps of the societal ladder.

6. Discussion

Jensen wrote in 1998 (p. 532) that the study of sex differences in general intelligence is " technically the most difficult to answer … the least investigated, the least written about, and, indeed, even the least often asked."

The purpose of the present study became accordingly to try and refine the methodological approach and to see whether unclear definitions and unreliable measures could explain why the area of sex differences had for so long been beleaguered by confusion, occasional glimpses of clarification, wildly differing interpretations, and hasty formation of conclusions, not rarely based more on "what ought to be" than on "what is" sexist attitudes.

Test of hypothesis 1 suggested that the disagreement among the experts could, indeed, be ascribed to conceptual and methodological problems. Test of hypothesis 2 confirmed that when intelligence is defined as general intelligence *g* and measured by HFA/SL analysis, a moderate average sex difference of .37 SD (or the IQ equivalent of 5.5 points) is identified, as is a larger male dispersion of *g* factor scores. Further analysis provided support for hypothesis 3, which states that the sex difference in *g* can be explained, at least in part, in terms of a larger average male brain. However, the considerable overlap of the male and female brain size distributions, as well as the moderate brain - *g* correlation, obviously temper this conclusion.

Classical Distribution Theory implies that the male average lead in *g*, and the larger male than female dispersion of scores, will manifest themselves in an exponentially growing male/female ratio at the high end of the distribution, with the exact numerical ratio being less a function of the size of the average sex difference in mean, than of the dispersion of *g*.

The present data can be used to enlighten the dynamic forces behind the worldwide massive male dominance in areas calling for very high general intelligence. Thus, assume a total population of 6 billions (half males, half females - not entirely correct, but this would favour females), and extrapolation from the present data identifies the most gifted female at $g = 4.559$ SD (IQ 168.379), and the most gifted male at $g = 6.904$ (IQ 203.563); above the most gifted female on earth there will be 111,975 males with higher *g*.

Obviously, no contemporary tests reliably taps these extreme *g* (or IQ) values, and the presently observed difference in dispersion (male SD = 1.06 versus female SD = .74) is larger than most literature on intelligence would lead us to expect. Then again, the present population is much more carefully sampled than usually, suggesting that the true sex difference in dispersion SD has so far been underestimated. The small N in the present study obviously calls for caution in interpretation, but it is harder to obtain a statistically significant difference in a small than in a large sample. Moreover, the Colom et al. (2002) study applied precisely the same analytic approach to a much larger sample (N = 1,369), and the male lead in *g* was statistically significant at the .0001 level.

To sum up, proper methodology identifies a male lead in general intelligence *g*, which relates to brain size and explains, in part, male dominance in society.

References

Aluja-Fabregat, A., Colom, R., Abad, F. & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as *g* among young adolescents. *Personality and Individual differences, 28(4)*, 813-820.

Anderson, B. (2003) Brain imaging and g. Chapter 2 in: H. Nyborg (ed.) *The scientific study of general intelligence: Tribute to Arthur R. Jensen.* Oxford: Pergamon.

Ankney, C. (1992). Sex difference in relative brain size: The mismeasure of woman, too? *Intelligence, 16,* 329-336.

Ankney, C. (1995). Sex difference in brain size and mental abilities: Comments on R. Lynn and Benbow, C. P. *Personality and Individual Differences, 18,* 423-424.

Asch, S. & Witkin, H. (1948). Studies in space orientation. I. Perception of the upright with displaced visual fields. *J.Exp.Psychol., 38,* 325-337.

Brody, N. (1992). *Intelligence* (2nd edition). San Diego, CA: Academic Press.

Colom, R., García, L.F., Juan-Espinosa, M. & Abad, F. (2002). Null sex differences in general intelligence: Evidence from the WAIS-III. *Spanish Journal of Psychology, 5(1)*, 29-35.

Colom, R. & García-López, O. (2002). Sex difference in fluid intelligence among high school graduates. *Personality and Individual Differences, 32*, 445-451.

Colom, R., Juan-Espinosa, M., Abad, F. & García, L. (2000). Negligible sex differences in general intelligence. *Intelligence, 28(1),* 57-68.

Christiansen, P.M., Møller, B., & Togeby, L. (2001) *Den danske elite* [The Danish Elite[. Copenhagen: Hans Reitzels Forlag.

Gignac, C., Vernon, P.A., & Wickett, J.C. (2003) Factors influencing the relationship between brain size and intelligence. Chapter 2 in H. Nyborg (ed.) *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. (pp. 93-106). Oxford: Pergamon.

Goldberg, S. (1977). *The inevitability of patriarchy*. London: Temple Smith.

Gottfredson, L. S. (2003). *g*, jobs and life. In Nyborg H. (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 293-342). Oxford, UK: Pergamon.

Haier, R.J. (2003) Positron emission tomography studies of intelligence: From psychometrics to neurobiology. Chapter 3 in: H. Nyborg (ed.) *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. Oxford: Pergamon.

Halpern, D.F. & LaMay, M.L (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12,2*, 229-246.

Jensen, A.R. (1980) *Bias in mental testing*. London: Methuen.

Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT., Praeger.

Jensen, A.R., & Sinha, S.N. (1993) Physical correlates of human intelligence. In P.A. Vernon (ed.) *Biological approaches to the study of human intelligence* (pp. 139-242). Norwood, NJ: Ablex.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences, 17*, 257-271.

Lynn, R. (1997). Sex differences in intelligence: Data from a Scottish standardization of the WAIS-R. *Personality and Individual Differences, 24,2*, 289-290.

Lynn, R. (1999). Sex difference in intelligence and brain size: A developmental theory. *Intelligence, 27,1*, 1-12.

Lynn, R., Irwing, P., & Cammock, T. (2002). Sex differences in general knowledge. *Intelligence, 30,1*, 27-39.

Money, J. (1965). *A standardized road-map test of direction sense*. Baltimore: The Johns Hopkins Press.

Nyborg H. (2001). Early sex differences in general and specific intelligence: Pitting biological against chronological age. (Addendum). Second Annual Conference of the International Society for Intelligence Research, Cleveland, OH.

Nyborg, H. (2002) IQ and *g*: The art of uncovering the sex difference in general intelligence. Paper presented at the Third Annual Conference of the International Society for Intelligence Research, Vanderbilt University, Nashville, TN, December 5-17th.

Nyborg, H. (2003) Sex differences in *g*. Chapter 10 in: H. Nyborg (ed.) *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. Oxford: Pergamon.

Nyborg, H. & Isaksen, B. (1974). A method for analysing performance in the rod-and-frame test. II. Test of the statistical model. *Scandinavian Journal of Psychology, 15,* 124-126.

Nyborg, H. & Jensen, A. R. (2001). Occupation and Income Related to Psychometric g. *Intelligence, 29,* 45-55.

Packenberg, B., & Gundersen, H.J.G. (1997) Neocortical neuron number in humans: Effect of sex and age. *The Journal of comparative Neurobiology, 384*, 312-320.

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

Rushton, J. P. & Ankney C.D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin and Review, 3*, 21-36.

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

Schoenemann, P.T., Budinger, T.F., Sarich, V.M., & Wang, W. (2000). Brain size does not predict general cognitive ability within families. *Proceedings of the National Academy of Sciences, 97*(9), 4932-4937.

Vandenberg, S. & Kuse, A. (1978). Mental rotations, a group test of three-dimensional visualization. *Perceptual and Motor Skills, 47,* 599-604

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence (4th. ed.).* Baltimore: Williams & Wilkins.

Witkin, H. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality, 51,* 1-15.

Figure 1. Male and female distributions and ratios as a function of male *g* = .36 (SD 1.06) and female *g* = -.01 (SD .74). (Data from Nyborg, 2003)

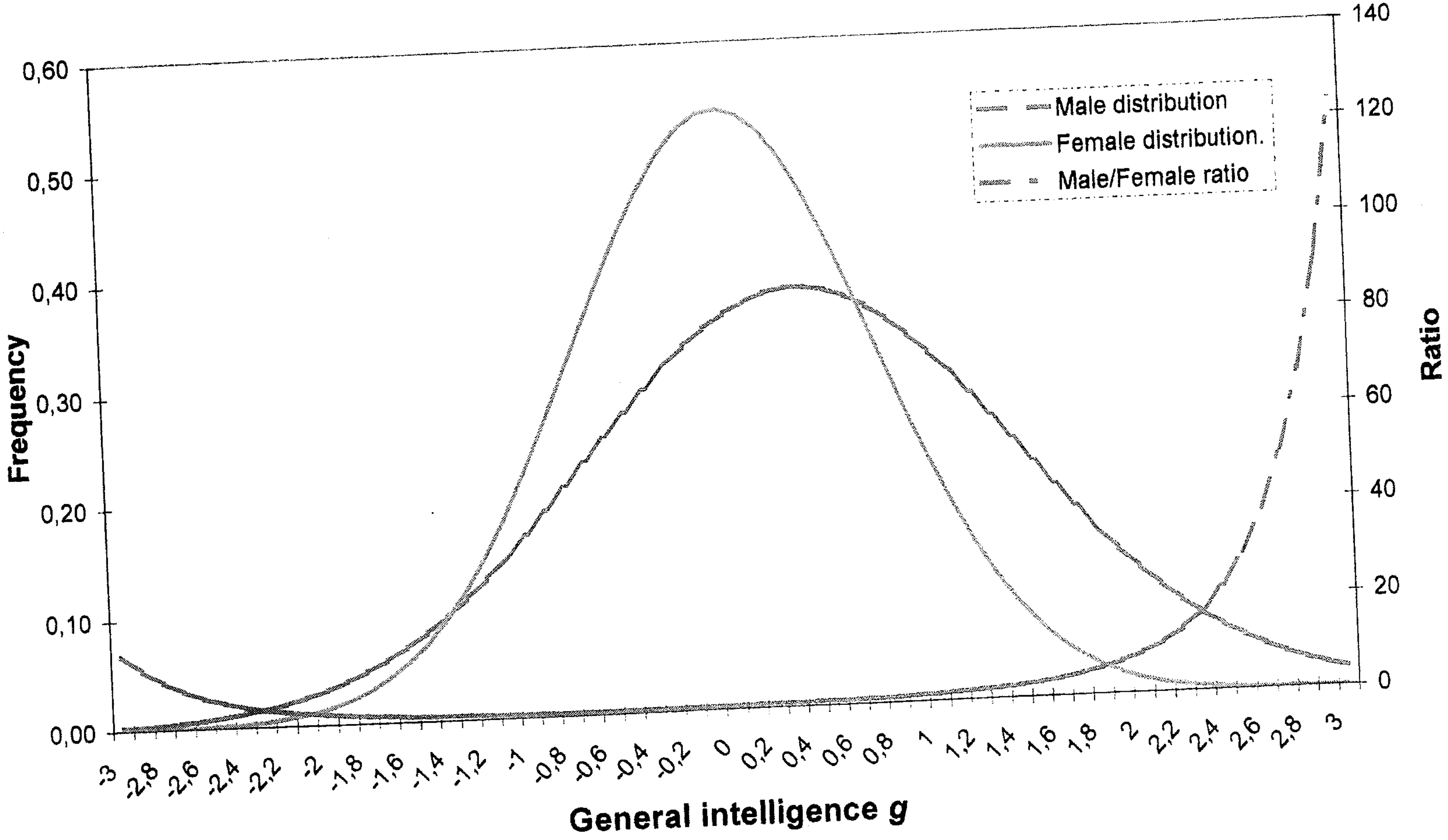

Table 1. Type of test, d effects, point-biserial correlations, and *g* loadings for 20 variables with eigenvalues > 1. Point-biserial correlations (r_{pbs} adjusted for unequal SDs) were factored in to reflect the loading of sex on the *g*-dimension (N = 26 males and 26 females).

Tests	**Effect** **d**	**Point-biserial correlation** r_{pbs}	***g* Loading Secondary factor**
RFT Frame dependence (signed error, inv.)	0.43	0.25	0.423
RFT Response variability (error, inv.)	0.42	0.25	0.424
RFT Field dependence (unsigned error inv.)	0.41	0.24	0.434
Embedded-Figures test (Sec/fig inv.)	0.13	0.07	0.466
Money left-right discrimination test (error inverted)	0.53	0.33	0.630
Mental Rotation (Figures found, corrected for guessing)	0.68	0.44	0.501
Tapping (Left hand)	0.60	0.38	0.241
Tapping (Right hand)	0.30	0.17	0.283
Oral fluency	-0.08	-0.04	0.182
WAIS Information	0.43	0.25	0.545
WAIS Comprehension	-0.22	-0.10	0.387
WAIS Arithmetic	0.13	0.07	0.469
WAIS Similarities	0.34	0.20	0.451
WAIS Digit Span	0.16	0.09	0.163
WAIS Vocabulary	0.35	0.20	0.487
WAIS Digit Symbol	-0.56	-0.19	0.025
WAIS Picture Completion	0.43	0.26	0.407
WAIS Block Design	0.08	0.04	0.668
WAIS Picture Arrangement	0.04	0.02	0.451
WAIS Object Assembly	-0.06	-0.03	0.535
Point-biserial factor loading of sex			**0.272***
Average Effect Size	0.26		
Average IQ equivalent	3.94		
Average factor loading			0.402

Notes:

* Significant at *p* = .0275, one-sided).

Table 2. Correlated vector analysis based on hierarchical *g* loadings and average sex differences in *g* factor scores (d) for 26 male and 26 female adult subjects on 20 tests, illustrating the extent to which the sex differences load on the *g* dimension. Reliability coefficients for the first ten tests are the lower-bound estimate of each test's communality. Reliablity coefficients for the WAIS sub-tests were taken from Jensen (1980).

Column	A	B	C	D	E	F	G	H	I	J
	Factor loadings					Mean sex differences				
	Uncorrected			*Corrected*		*Uncorrected*			*Corrected*	
Test	*g*	Rank	Reliability	*g*	*Rank*	d	Rank	Reliability	d	*Rank*
RFT Frame dependence (signed error, inv.)	.423	8	.867	.454	6	.427	15	.867	.458	14
RFT Response variability (error, inv.)	.424	9	.658	.523	10	.419	14	.658	.517	16
RFT Field dependence (unsigned error, inv.)	.434	10	.892	.460	7	.408	13	.892	.432	13
Embedded-Figures test (sec/fig inv.)	.466	13	.475	.676	17	.127	8	.475	.184	8
Money left-right discrimination test (Inv.)	.630	19	.785	.711	18	.533	18	.785	.601	17
Mental Rotation (Figures found, corrected f. guess.)	.501	16	.586	.654	15	.676	20	.586	.882	20
Tapping (Left hand)	.241	4	.732	.281	4	.599	19	.732	.700	18
Tapping (Right hand)	.283	5	.634	.355	5	.301	10	.634	.378	10
Oral fluency	.182	3	.675	.222	3	-.080	3	.675	-.097	3
WAIS Information	.545	18	.718	.644	14	.427	16	.718	.504	15
WAIS Comprehension	.387	6	.596	.501	8	-.220	2	.596	-.285	2
WAIS Arithmetic	.469	14	.738	.546	11	.126	7	.738	.147	7
WAIS Similarities	.451	12	.795	.506	9	.345	11	.795	.387	11
WAIS Digit Span	.163	2	.664	.200	2	.159	9	.664	.196	9
WAIS Vocabulary	.487	15	.730	.570	12	.352	12	.730	.412	12
WAIS Digit Symbol	.025	1	.518	.035	1	-.562	1	.518	-.780	1
WAIS Picture Completion	.407	7	.318	.721	19	.428	17	.318	.758	19
WAIS Block Design	.668	20	.747	.773	20	.084	6	.747	.098	6
WAIS Picture Arrangement	.451	11	.590	.587	13	.039	5	.590	.051	5
WAIS Object Assembly	.535	17	.660	.659	16	-.062	4	.660	-.076	

Correlations (one-tailed)	A with F	D with I
Pearson *r*	.431, *p* = .029	.455, *p* = .022
Spearman rank-order r_s	.245, *p* = .149	.268, *p* = .127

Table 3. Prediction of sex differences in intelligence in general IQ or general intelligence *g* from observed sex differences in scanned brain Volume, or from head circumference, which is a rough proxy for brain size. (re-calculated from Nyborg , 2001; 2002; 2003).

Study	A: Observed sex difference in brain volume[a] or circumference[b]	B: Correlation between volume [a] or circumference[b] and IQ[1] or *g* factor score[2]	C: A x B	D: Predicted male lead in IQ[3] or *g* factor score[4]	E: Observed male lead in IQ[5], *g* factor scores[6], or d[7] (based on sex loaded r_{pbs})
	d	*r*	**SD units**	**C x 15**	
Intelligence in general IQ data (Lynn, 1999, average over several studies)	.78[a]	.35[a][1]	.27	4.05[3]	3.85[5]
General intelligence *g* data (Nyborg, 2001; 2002; 2003, fig. 10.2)	.87[b]	.34[b][2]	.30	.30[4] *(IQ 4.44)*	.37[6]*(IQ 5.55)*, or .57[7]*(IQ 8.55)*

Note:
the *g* factor (.272, see table 1), the formula given by
[7] After the r_{pbs} was factored in on the *g* dimension to indicate how much sex loads on the *g* factor (.272, see table 1), the formula given by Jensen (1998, p. 543, note 12) was used to derive its d value (see table 2), which, when multiplied by 15 provides its IQ equivalent (in brackets).