102. (a) Lucas et al., 1992. (b) Ryan et al., 1991. Leary, 1988.
103. (a) Eysenck, 1991. (b) Eysenck, 1995; Eysenck & Schoenthaler, 1997.
104. (a) Olson (1994) provides a review and extensive bibliography on the psychological effects of fetal alcohol syndrome. (b) Reinisch et al., 1995.
105. Snyderman & Rothman, 1988, pp. 128–130, 294.
106. (a) Ogbu (1978) is the most comprehensive presentation of this view. (b) Ogbu (1994) summarizes his views on cultural determinants of intelligence differences.
107. Sowell, 1981, 1994; Chapter 6 specifically deals with race and mental ability.
108. The conspicuous success of Asian immigrants in intellectual pursuits in America is extensively documented in P. E. Vernon (1982), Flynn (1991), Caplan et al. (1992).
109. Spearman & Jones, 1950, p. 72.
110. Steele & Aronson, 1995; Steele, 1997.
111. (a) Spielberger & Sydeman (1994) provide a review and bibliography of research on test anxiety. See also Jensen, 1980a, pp. 615–616. (b) Yerkes & Dodson, 1908. (c) Spielberger, 1958.
112. Lovejoy, 1993.

The g factor : The science of mental abili
Westport, CT: Praeger
Bilag A(0)

Chapter 13

Sex Differences in *g*

Past studies of a sex difference in general ability have often been confounded by improper definitions and measurements of "general ability" based on simple summation of subtest scores from a variety of batteries that differ in their group factors, by the use of unrepresentative groups selected from limited segments of the normal distribution of abilities, and by the interaction of sex differences with age-group differences in subtest performance. These conditions often yield a mean sex difference in the total score, but such results, in principle, are actually arbitrary, of limited generality, and are therefore of little scientific interest. The observed differences are typically small, inconsistent in direction across different batteries, and, in above-average samples, usually favor males.

In this chapter sex differences are specifically examined in terms of their loadings on the *g* factor for a number of test batteries administered to representative population samples. When the sex differences (expressed as a point-biserial correlation between sex and scores on each of a number of subtests) were included in the correlation matrix along with the various subtests and the correlation matrix was subjected to a common factor analysis, sex had negligible and inconsequential loading on the *g* factor, averaging about .01 over five test batteries. Applying the method of correlated vectors to these data shows that the magnitude of the sex difference on various subtests is unrelated to the tests' *g* loadings. Also, the male/female variance ratio on diverse subtests (generally indicating greater male variability in scores) is unrelated to the subtests' *g* loadings. Although no evidence was found for sex differences in the mean level of *g* or in the variability of *g*, there is clear evidence of marked sex differences in certain group factors and in test specificity. Males, on average, excel on some factors; females on others. The largest and most consistent sex difference is found on a spatial-visualization factor that has its major factor loadings on tests requiring the mental rotation or manipulation of figures in an imaginary three-dimensional space. The difference is in favor of males and within each sex is related to testosterone level. But the best available evidence fails to show a sex difference in *g*.

Research on sex[1] differences in mental abilities has generated hundreds of articles in the psychological literature, with the number of studies and articles increasing at an accelerating rate in the last decade. As there now exist many general reviews of this literature,[2] I will focus here on what has proved to be the most problematic question in this field: whether, on average, males and females differ in *g*.

It is noteworthy that this question, which is technically the most difficult to answer, has been the least investigated, the least written about, and, indeed, even the least often asked.

The vast majority of studies have looked at sex differences in more specialized abilities, such as can be subsumed under the labels of certain well-established primary (first-order) or group factors in the psychometric domain. In the three-stratum hierarchy of ability factors, sex differences also appear at the second stratum.

The differences observed for specific tests and for first-order and second-order factors are now well established by countless studies. They constitute an empirical fact and the frontier of research now lies in discovering the causes of the clearly identified cognitive differences between the sexes. However, a brief examination of these first-order psychometric differences is necessary in order to understand the problem of determining whether the sexes differ in *g*.

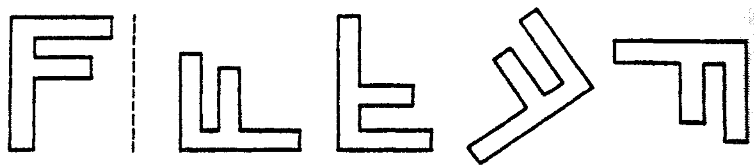
SEX DIFFERENCES IN SPECIFIC TESTS AND IN FIRST-ORDER FACTORS

Neither Binet, in the development of his test, nor Terman, in creating the American version and the initial standardization of Binet's test (known as the Stanford-Binet), took account of sex differences. Because a sex difference in the overall test scores was of negligible size (although a minority of the individual items showed sex differences, some favoring girls and others favoring boys, and these differences at the item level largely averaged out in the composite score), it was assumed that the sexes did not differ significantly in what the test as a whole was intended to measure, namely, general intelligence. Therefore, in all subsequent revisions of the Stanford-Binet (including the latest revision, the Stanford-Binet IV) any items that showed exceptionally large and statistically significant sex differences were eliminated in order to counterbalance the sex differences for the remaining items. This counterbalancing is possible, of course, only if roughly equal numbers of sexually discriminating items favor each sex, and these discriminating items must be found throughout the range of item difficulty. Given these built-in features, obviously, the Stanford-Binet scale is hardly a suitable research instrument for studying sex differences. The negligible sex difference in Stanford-Binet IQ could be just an artifact of item selection.

A number of other standardized tests were similarly designed to minimize sex differences, the best known being the Wechsler Intelligence Scales. Because the Wechsler test consists of a dozen distinctive subscales, half of them verbal and half nonverbal (or performance), it is not easy, or perhaps not even possible, to equate the sexes on each of the subscales. These tap not only *g*, but also certain first-order factors (particularly verbal, spatial, and short-term memory). If the sexes actually differ on subtests that are aggregations of items that mainly reflect different first-order factors, the overall difference on a given subtest cannot be eliminated by item selection and still be differentiated factorially from the other subtests in the battery. This is in fact the case with the Wechsler scales. Even though the few most sexually discriminating items have been eliminated from each subtest, males and females still differ significantly on the various subtests. Females slightly but significantly exceed males on some verbal tests and on short-term memory; males exceed females on some of the performance tests, particular those calling for spatial visualization. The fact that these differences largely balance out in the overall IQ score could be a result of selection of the types of subscales used in the Wechsler, whether intentional or inadvertent. So the Wechsler IQ per se permits no conclusion about a sex difference in intelligence. Because sex differences were taken into account in the construction of the Wechsler scales, we can infer sex differences in the special abilities measured separately by each of the subtests but cannot convincingly estimate their magnitudes. For that we must turn to tests that were constructed entirely without reference to sex.

In these studies the sex difference is typically measured in terms of standard deviation units, termed the "effect size" or the *d* statistic. The value *d* is defined as the quantity male mean minus female mean divided by the average within-group standard deviation, i.e., $d = (\overline{X}_M - \overline{X}_F)/\overline{\sigma}$.

Visual-Spatial Abilities. These abilities favor males and have the largest and most consistent sex differences of any psychometric abilities. The factor analysis of various kinds of tests of visual-spatial abilities reveals about ten distinct subfactors.[3] All of these analyses show a sex difference favoring males, but the largest difference is on tests that require visualizing 2-dimensional or 3-dimensional figures that have to be rotated or manipulated mentally in 3-dimensional space. For example, the testee must determine which of the following four figures on the right are the same as the figure on the left flipped over and rotated rather than merely rotated in its own plane? (*Answer*: b and c.)



a b c d

It is important to note that not every test of figural material involves the spatial ability factor. Raven's Progressive Matrices, for example, does not qualify as a spatial test. This is shown by the fact that when the Raven is factor-analyzed among a variety of tests including several tests that define a spatial factor, it does not significantly load on the spatial factor. The defining characteristic for spatial problems is that, in order to obtain the correct solution, the subject must visualize and manipulate the figural material mentally as if it were an object in three-dimensional space. Men, on average, excel women in this type of performance. Meta-analyses of the sex difference on various composites of spatial-visualization tests yield average *d* values in the range of .30 to.50 for the general population.[2c,d,i,4]

The only sex differences favoring males that are larger than this do not involve broad factors as such, but occur on tests in which spatial ability is combined with types of specific knowledge content with which males are typically more familiar (such as information about tools, auto mechanics, and electronics). Tests involving such knowledge were developed to measure vocational aptitudes for selecting individuals, usually men, for certain skilled jobs or job training. On some of these tests males exceed females by a *d* of about 1.0 ±.3. Such large sex differences occur only on tests that reflect specific achievement rather than the broad abilities that emerge at the second order in factor analyses.

Since the sex difference in spatial ability is ubiquitous throughout the human species (and is even found in other mammalian species), the consensus of expert opinion today doubts that the phenomenon is explainable purely in terms of environmental or cultural factors. The sex difference in spatial ability appears to be a *sex-limited* trait, which means that the genetic basis of individual differences in the trait is the same for both sexes, but that some other factor that differs between the sexes has the effect of limiting the expression of the trait in one sex. The best present evidence is that this additional factor involves the individual's estrogen/testosterone balance, which of course differs quite markedly in men and women.[5] Within each sex there is a nonlinear (inverted-U) relationship between an individual's position on the estrogen/testosterone continuum and the individual's level of spatial ability, with the optimal level of testosterone above the female mean and below the male mean. Generally, females with markedly above-average testosterone levels (for females) and males with below-average levels of testosterone (for males) tend to have higher levels of spatial ability, relative to the average spatial ability for their own sex. Other hypotheses based on sexual dimorphism in certain brain structures (particularly the corpus callosum) and sexual differences in the development of hemispheric dominance are also being considered, as is the evolutionary basis of these differences.[6]

Mathematical Reasoning Ability. Because mathematical or quantitative reasoning is a prominent feature of many scholastic and employment aptitude tests, including the most widely used of all such tests (the SAT), the repeated finding of a rather marked sex difference in this aptitude has given rise to a great amount of research. In recent years, this topic has even become a prominent research specialty of behavioral scientists.[7]

The sex difference favoring males does not include ability in arithmetic calculation, in which females slightly excel males, but exists for quantitative "thought problems" and especially for the more advanced and complex aspects of mathematics taught in high school and college. The sex difference in math ability in the general population is not large, with *d* values mostly between .10 and .25 for various tests given to nationally representative samples. Much larger differences appear in subject samples that were selected from the upper-half of the distribution of math ability; the further above the general population mean, the larger the sex difference. One reason for this is the considerably greater variance of males in math ability. The variance of males' math test scores averages about 1.1 to 1.3 times greater than the variance for females. Almost twice as many males as females fall into the upper tail (>90th percentile) of the bell-curve distribution of math scores. However, males also outnumber females in the lower tail (<10th percentile) of the math score distribution. This phenomenon of greater male variance, which is most conspicuous in the extreme tails of the distribution, is generally found for most psychometric abilities. But it is more extreme in both math and spatial abilities than in any other broad ability factors. Data collected between 1970 and 1990 or so suggest that there has been a slight decrease in the sex difference in math ability in representative population samples.

Causal theories of the math sex difference are still tentative and debatable, pending further investigation. Some theorists attribute the math difference to the somewhat larger sex difference in spatial ability, in part because some types of math problems can be visualized graphically or in terms of spatial relations. But the moderate correlation of math ability with spatial ability (independent of *g*) is generally too small to account for more than a minor fraction of the sex difference in math ability. Space and math have independent determinants besides the source of variance they share in common. Biological and evolutionary psychologists have proposed theories similar to those for spatial ability, explaining the well-substantiated sex difference in math ability in terms of natural selection for the different roles performed by males and females in the course of hominid evolution and their genetically transmitted neurophysiological and hormonal mediators.[7a,b]

Verbal Abilities. The sex difference on verbal tests for young adults fluctuates around zero across various tests and studies, and seems to depend more on specific properties of each verbal test rather than reflecting any consistent difference. Girls show more rapid maturation than boys in verbal expression, but this difference begins to disappear after puberty. In general, verbal tests with the largest *g* loadings, such as tests of verbal reasoning, show differences averaging close to zero. It is possible to devise tests that use a verbal medium but that emphasize abstract reasoning more than verbal knowledge or fluency. The one type of verbal ability that most consistently favors females is *verbal fluency*. A typical test requires the subject to produce as many common nouns beginning with a given letter within a limited time (say, thirty seconds). Scholastic-type achievement tests involving verbal content, such as reading, writing, grammar, and spelling, also consistently favor females. Tests of general information and especially science information and technical knowledge favor males.

Smaller Group Factors. Perceptual speed and short-term memory both favor females, with d values of $-.20$ to $-.30$. One of the larger sex differences favoring females is on a factor identified as "speed and accuracy" or "clerical checking." This factor is measured by the digit symbol or coding test of the Wechsler battery. Typical d values fall between $-.30$ and $-.40$. A sex difference on a test of this factor as high as $-.86$ was found for male and female twelfth-grade high school students taking the General Aptitude Test Battery's subtest Q ("clerical perception"), which makes a great demand on perceptual speed and accuracy.

THE *g* FACTOR: A NEAR-ZERO SEX DIFFERENCE

The study of a sex difference in general ability has often been confused by the use of different concepts of "general ability," and by failing to recognize that the typically greater variance of males in test scores may cause both the direction and magnitude of the mean sex differences in test scores to vary across different segments of the total distribution for the general population. The observed sex difference will therefore often vary across groups selected from different segments of the population distribution.

Many investigators have taken the sum total of the subtest scores on one or another IQ test as their operational definition of general mental ability. The total IQ is usually based on the sum of the standardized scores on the various subtests. Because there are sex differences on various types of subtests, as previously noted, the direction and magnitude of the summed sex differences will depend on the particular selection of subtests in the battery. For example, including more tests loaded on the spatial factor will favor males; including more tests loaded on verbal fluency and on clerical speed and accuracy will favor females. Thus the simple sum or mean of various subtest scores is a datum without scientific interest or generality. It cannot be considered a proper measure of general ability. If, for a particular battery, it happened to be an adequate measure of general ability, it would be so only inadvertently, as a result of averaging out sex differences that have more or less equal and opposite signs on the various subtests. This becomes more likely the greater the number of different types of subtests averaged. The essential problem is that the concept of general ability, defined as *g*, rests on the *correlations* among test scores rather than on their simple summation. The latter, which might be referred to as "ability in general," is an arbitrary variable that fails to qualify conceptually as a scientific construct, although by happenstance it may correlate highly with general ability defined as *g*.

But even if there are many subtests in a battery, thereby tending to average out sex biases, a simple summation of sex differences over subtests is contaminated if the method of test construction included selecting test items on any criteria involving sex differences in item responses (as was done in creating the Stanford-Binet and the Wechsler intelligence scales). IQ scores on such tests can hardly be informative about the magnitude of possible sex differences in general ability, at least in principle. The study of sex differences must depend on tests in which item selection was based exclusively on the psychometric criteria used to maximize the reliability, validity, discriminability, and unidimensionality of the subtests.

Studies that are based on selected samples of males and females are highly questionable if the results are generalized to any population other than the one from which the study sample was selected. Because males' scores are more variable on most tests than are those of females, there are more males at both the upper and the lower extremes of the distribution. This is most markedly true for tests of spatial and quantitative abilities. Many of the published studies of sex differences have been based on self-selected or institutionally selected groups that score mostly in the upper half or even the upper quarter of the normal distribution of abilities, groups such as college applicants, university students, and people in highly skilled occupations, such as Air Force officers. Generally, the higher the cut-score for selection, the larger is the proportion of exceptionally high-scoring males in the group. The mean sex difference in such an above-average group would not accurately estimate the magnitude of the sex difference in the general population, but would yield a biased estimate favoring males. The opposite bias would appear if the sexes were compared in a group that scored well below the general population mean. Although greater male variance is typically found in American and north European studies, this phenomenon is not generalizable across all nations and cultures.[8] A study of sex difference in general ability intended to be generalizable to recent North American or north European populations should be based on representative samples of those populations. There are few such data sets available after excluding the national standardization samples for tests in which item selection took sex into account.

The age of the study sample must also be taken into account. Girls mature earlier than boys, which favors girls' language development and verbal facility in childhood. Also, sex differences in spatial and quantitative abilities, which are relatively nascent in childhood, are affected by the hormonal changes after puberty, although some part of the effect of testosterone on spatial ability occurs prenatally. The study of asymptotic sex differences, therefore, is focused on representative samples in adolescence and early maturity. In later maturity and old age, sex differences in health factors and in longevity interact with sex differences in cognitive abilities, limiting the generalizability of the findings. The following studies of sex differences in *g* are based largely on representative population samples of individuals in adolescence and early maturity, except for one test on children and adolescents (WISC-R, ages 5 to 16) that is included for comparison with a parallel test for young adults (WAIS, ages 23 to 34).

Factor Analyzing Sex Differences. The best method for determining the sex difference in psychometric *g* is to represent the sex difference on each of the subtests of a battery in terms of a point-biserial correlation and include these correlations with the full matrix of subtest intercorrelations for factor analysis.[9] The results of the analysis will reveal the factor loading of sex on each of the factors that emerges from the analysis, including *g*. The *g* factor loading of sex, therefore, is equivalent to the point-biserial correlation between *g* and the sex variable (quantitized as male = 1, female = 0). This method is preferable to the use of *g* factor scores (which I used in an earlier study[10] of sex differences on the WISC-R), because *g* factor scores are not a pure measure of the *g* factor of the test battery from which it was extracted. An individual's *g* factor score is calculated as a *g*-weighted mean of the individual's standardized scores on each of the subtests; it is therefore necessarily somewhat contaminated by including small bits of the other factors and test specificity measured by the various subtests. This contamination of *g* factor scores can either increase or decrease the mean sex difference, depending on the types of subtests in the battery. Therefore, it is better to factor analyze the matrix of all the subtest intercorrelations, including the correlations of sex with each of the subtests.

I have performed this analysis with five test batteries in which data were available for large and representative samples that encompass the full range of ability in the general population. The results are shown in Table 13.1. Its interpretation requires some information about the test batteries and the key variables derived from them.

For all of the test batteries, the subject samples were expressly selected to be representative of the stated age groups in the general population. For every test there were either equal or very nearly equal percentages of males and females, the largest difference (GATB) being only 1.8 percent.

The WISC-R and the WAIS are the child and adult versions of the Wechsler Intelligence Scales; the data are from the national standardization samples.[11a] The various subtests are of the same types in both batteries, but differ in the level of difficulty. Although items (within each subtest) that showed marked sex differences were eliminated in the construction of the Wechsler tests, neither the items nor the subtests were selected with any reference to factor analysis. Whatever relationship emerges between sex differences and the factor composition

Table 13.1
Relationship of Sex Differences to *g* in Five Test Batteries

Variable	Test Battery[a] WISC-R	WAIS	GATB	ASVAB	BAS	Mean	Median
Number of Subtests	13	11	8	10	14		
g Loading of Sex Difference[b]	.094	.006	-.255	.180	-.001	.011	.006
d Equivalent of *g* Loading	.189	.012	-.527	.366	-.002	.008	.012
IQ Equivalent of d	2.83	0.18	-7.91	5.49	-0.03	0.11	0.18
Percent of Total *g* Variance Due to Sex Differences	0.19	0.00	2.27	0.54	0.00	0.60	0.19
Correlation (r_s) of Tests' *g* Loadings With Sex Differences	.364	-.036	.024	.127	.103	.116	.103
Correlation (r_s) of Tests' *g* Loadings With M/F Variance Ratios	-.261	.318	-.738	.079	-.033	-.127	-.033

[a]Wechsler Intelligence Scale for Children-Revised (WISC-R), U.S. standardization sample, ages 5 to 16 years; Wechsler Adult Intelligence Scale (WAIS), American standardization sample, ages 25 to 34 years; General Aptitude Test Battery (GATB), unselected twelfth-grade high school students; Armed Services Vocational Aptitude Battery (ASVAB), a nationally representative sample of American youths, ages 18 to 23 years; British Ability Scales (BAS), British standardization sample, ages 14 to 17 years.
[b]Positive and negative loadings indicate male superiority and female superiority, respectively.

and relative magnitudes of factor loadings of the Wechsler tests, therefore, is not an artifact of the method of test construction.

The GATB analysis was based on the normative data for twelfth-grade students in high school.[11b] Sex differences were not considered in the construction of the GATB.

The ASVAB data are based on a large probability sample of the U.S. youth population.[11c] The mean sex differences used in this analysis were based on only non-Hispanic whites. Sex differences did not enter into the construction of the ASVAB, and the various subtests show very marked sex differences, especially on subtests involving technical information to which men are generally more exposed (e.g., auto shop, mechanical reasoning, electronic information).[11d] However, the factor structure per se of the 10 ASVAB subtests combined with 17 psychomotor tests is the same for adult males and females and the *g* loadings of the 27 tests are correlated +.999 between males and females.[11f]

From a strictly psychometric standpoint, the British Ability Scales (BAS) are probably as well constructed a battery of cognitive ability tests as one can find at present. Each of the fourteen subtests was constructed by means of item selection based on item response theory, or a latent trait model (in this case the Rasch model), which, for this type of test, is the optimal psychometric procedure. Test construction is based entirely on procedures that maximize psychometric desiderata without reference to sex or other subclassifications of the normative population. The normative samples for ages fourteen to seventeen were used in the present analysis.

Most of the variables listed in Table 13.1 are self-evident: The g loading of the sex difference is equivalent to the point-biserial correlation of sex with the test battery's g factor (here represented by the first principal factor of a common factor analysis).

The d equivalent of the g loading represents the size of the mean sex difference on g expressed in standard deviation units.[12]

The sex difference in IQ units is simply $15d$.

The percent of the total g variance due to sex differences is the squared g loading for sex divided by the total variance of the g factor (excluding sex variance) $\times$ 100.

The Spearman rank-order correlation (r_s) of the column vector of subtests' g loadings with the vector of the sex differences (d) on the subtests indicates the degree to which g is related to the rank order of the sex differences on the various subtests.

The method of correlated vectors was also applied to the vector of the M/F variance ratios for the various subtests, which measures the degree to which males are more variable than females. For all but one test (WAIS), greater male variability on the subtests is *negatively* correlated with the subtests' g loadings.

Several points are especially worth noting:

- The g loadings of the sex differences are all quite small; the largest difference (GATB) favors females (−.255). (The GATB results are somewhat discrepant from those of the other batteries because of the rather unusual psychometric sampling, with an excess of psychomotor tests designed to measure particular vocational aptitudes. This slightly compromises the g factor by diminishing its loadings on the more cognitive variables.[13]) The BAS, probably the best constructed of all of the tests from a psychometric standpoint, shows a near-zero g loading (−.001) on sex. The mean and median g loadings over the five tests are near-zero and completely nonsignificant.
- The method of correlated vectors shows that in no case is there a correlation even approaching significance between subtests' g loadings and the mean sex differences on the various subtests.
- The method of correlated vectors shows mostly negative (but nonsignificant, except for the GATB) correlations between the M/F variance ratios on each of the subtests and the subtests' g loadings.

The two main conclusions supported by the analyses in Table 13.1:

- The sex difference in psychometric g is either totally nonexistent or is of uncertain direction and of inconsequential magnitude.

- The generally observed sex difference in variability of tests scores is attributable to factors other than g.

Consistent with this finding of a near-zero sex difference in g is the fact that there is no consistent sex difference on Raven's Standard Progressive Matrices (SPM) test (for adults) or on the Colored Progressive Matrices (CPM) test (for children). In numerous factor analyses, the Raven tests, when compared with many others, have the highest g loading and the lowest loadings on any of the group factors. The total variance of Raven scores in fact comprises virtually nothing besides g and random measurement error. In a review of the entire literature (117 studies from five continents) reporting sex differences on the Raven tests, Court[14] found positive and negative mean differences in the various studies distributed about equally around zero, for both the SPM and the CPM. Court concluded that there is no consistent evidence of a sex difference in the Progressive Matrices and that the most common finding is of no significant sex difference.

Some of the recent research on sex differences in IQ has been prompted by the finding of a significant sex difference in brain size, even when body size is statistically controlled (see Chapter 6, p. 148). The fact that brain size is correlated with IQ, and particularly with g, would seem to make the absence of a sex difference enigmatic. In an attempt to resolve this paradox, Lynn[15] has argued that a difference of only four IQ points, favoring males, would be consistent with the prediction from the sex difference in brain size and the within-sex regression of IQ on brain size.

In a review of several tests in several population samples, he has found an overall sex difference of about four IQ points. For the reasons pointed out above, any small overall difference (even if significant) on an arbitrary collection of subtests has questionable generality across different batteries and, in principle, cannot answer the question concerning a sex difference in general ability defined as g. Moreover, the sex difference in brain size may be best explained in terms of the greater "packing density" of neurons in the female brain, a sexual dimorphism that allows the same number of neurons in the male and female brains despite their difference in gross size. Also, the relationship of brain size to the well-established sex difference in spatial ability (independent of g) has not yet been studied. But if any feature of sex differences in brain anatomy or physiology is likely to be related to cognitive abilities, spatial ability is the best bet. In any case, variation in total brain size (or in whatever causes it, such as the number of neurons, arborization of dendrites, amount of myelin, etc.) accounts for only a minor part the total variance in g or in IQ. Other physiological factors unrelated to brain size, in addition to certain experiential factors, must also contribute a large part of the g variance.

The theoretical importance of finding a negligible sex difference in g is that it suggests that the true sex differences reside in the modular aspects of brain functioning rather than in whatever general conditions of the brain's information-processing capacity cause positive correlations among all of the modular functions on which there is normal variation and which account for the existence of *g*.

NOTES

1. Much of the recent literature on sex differences is unfortunately indexed and catalogued under the heading of *gender* differences, which is clearly inappropriate terminology for the topic of sex differences, as will be readily perceived by anyone who looks up the meaning of *gender* in an unabridged dictionary. A *sex* difference is any statistically significant difference in a characteristic between groups of individuals who posses the XY (male) and those who possess the XX (female) chromosome pairs.

2. Some key references on sex differences in mental abilities: (a) Brody, 1992, pp. 317–328; (b) Feingold, 1993, (c) Halpern, 1992; (d) Hedges & Nowell, 1995; (e) Hyde, 1981; (f) Jensen, 1980a, Chapter 13; (g) Kimura & Hampson, 1993; (h) Maccoby & Jacklin, 1974; (i) Mackintosh, 1996; (j) Stumpf, 1995.

3. Lohman's (1988) article on the nature of spatial abilities is the best treatment I have found of this topic.

4. Lubinski & Humphreys, 1990, Table 3.

5. Nyborg, 1984; Halpern, 1992, pp. 110–135; Kimura & Hampson, 1993; Feingold, 1996.

6. McKeever, 1995; Geary, 1995.

7. (a) Benbow, 1988; (b) Geary, 1996; (c) Lubinski & Humphreys, 1990. The references and peer commentaries for these key articles provide a fairly comprehensive bibliography of the modern research on sex differences in mathematical ability.

8. Feingold (1994) reviewed cross-national and cross-cultural differences in the variability of males and females on cognitive tests, concluding, "Cross-cultural comparisons . . . revealed large fluctuations in sex differences [in variability] across samples from diverse countries, suggesting that cultural factors are implicated in the results found in American samples" (p. 81).

9. The point-biserial correlation (r_{pbs}) is simply a Pearson product-moment correlation that expresses the relationship between a metric variable (e.g., test scores) and a dichotomous variable (in this case sex, quantitized as male = 1, female = 0). As the value of r_{pbs} is reduced by the amount of inequality in the sample sizes of males and females, it was corrected for this inequality where such an inequality in *N*s exists. Also, as r_{pbs} is reduced by an inequality of male and female standard deviations in test scores, the r_{pbs} was adjusted accordingly. Adjustments for the inequality of *N*s and *SD*s are accomplished simultaneously by use of the following formula for r_{pbs}:

$$r_{pbs} = d/2\sqrt{(d^2/4) + 1},$$

where d is the mean difference (males − females) divided by the averaged male and female standard deviations ($\bar{\sigma}$), calculated as $\bar{\sigma} = \sqrt{\sigma_m^2 + \sigma_f^2)/2}$.

Including the sex r_{pbs} for each of the subtests in the correlation matrix to be factor analyzed had no effect on the factor structure and only a negligible effect on the subtests' *g* loadings (congruence coefficients for all batteries are .999) when the factor analyses that include r_{pbs} in the correlation matrix were compared with the analyses that excluded r_{pbs} from the matrix. Therefore, it was not necessary to perform a Dwyer (1937) extension analysis (a mathematical maneuver that would be used in this case to isolate the sex variable itself from influencing the psychometric factors while showing its loadings on the psychometric factors).

10. Jensen & Reynolds (1983, Table 3) found a sex difference (M-F) of $d = +.161$ in *g* factor scores obtained from the national standardization sample (ages five to sixteen years) of the Wechsler Intelligence Scale for Children-Revised (WISC-R).

11. (a) The WISC-R and WAIS data (separately for males and females) were obtained from the publishers of the Wechsler tests. Factor analyses of the data separately by sex show the same factor structure for both sexes; the male × female congruence coefficients on the *g* factor are .99+; in the present analyses the within-sex *g* factor loadings were averaged. (b) The GATB data for high school seniors are from Tables 20-3 and 20-20 in the GATB Manual (U.S. Dept. of Labor, 1970). (c) From the Office of the Assistant Secretary of Defense, 1982, pp. 65 and 77. (d) A detailed discussion of sex differences on each of the ten ASVAB subtests and their occupational implications is provided by Bock & Moore, 1986, pp. 114–148. (e) From the Technical Handbook of the *British Ability Scales* (Elliott, 1983, pp. 63–88 and p. 152). (f) Carretta & Ree, 1997.

12. The *g* loading (or any point-biserial correlation, r_{pbs}, with equal N_s and equal SD_s of the dichotomous groups on the metric variable) is converted to d by the formula:

$$d = \sqrt{4/[(1/r_{pbs}^2) - 1]}.$$

13. Besides containing subtests of verbal (V), numerical (N), and spatial (S) abilities, which are common to many other test batteries intended to measure general cognitive ability, the GATB also contains a number of subtests intended to measure specific vocational aptitudes that strongly involve perceptual-motor abilities, such as perceptual speed of matching figures (P), clerical speed and accuracy (Q), motor coordination (K), finger dexterity (F), and manual dexterity (M). (Females slightly exceed males to varying degrees on V, P, Q, K, and F.) The greater number of perceptual-motor subtests in the GATB causes its first principal factor (PF1) to differ considerably from that of any of the other batteries, in which strictly cognitive abilities are relatively more represented than they are in the GATB. A factor analysis of the correlation matrix including the sex difference with just the cognitive variables (i.e., V, N, S) shows a sex loading on the general factor (PF1) of +.021, which is negligible. The general factor of the five perceptual-motor variables (P, Q, K, F, M) shows a sex loading of −.329, in favor of females.

14. Court, 1983.

15. Lynn (1994b), in a review of the sex differences (M-F) on Wechsler IQ (WISC-R and WAIS) obtained in several countries, reported Full Scale IQ differences ranging from 1.0 to 5.0 IQ points, with a mean of 3.08. Lynn's review also includes other tests, which show an average M-F difference of about four IQ points. He noted that the four IQ points male advantage is "precisely the advantage that can be predicted from their larger brains" (p. 269). His prediction is based on a reported correlation of .35 between *in vivo* brain size (measured by magnetic resonance imaging) and WAIS IQ and a reported sex difference of 0.78σ in adult brain size (based on autopsied brains), hence a predicted M-F difference in IQ of .35 × 0.78σ = 0.27σ × 15 ≈ 4 IQ points.