Available online at www.sciencedirect.com

ScienceDirect

Intelligence 35 (2007) 23 – 39

INTELLIGENCE

Sex differences in mental abilities: *g* masks the dimensions on which they lie

Wendy Johnson *, Thomas J. Bouchard Jr.

Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, USA

Received 8 June 2005; received in revised form 31 January 2006; accepted 9 March 2006
Available online 19 May 2006

Abstract

Empirical data suggest that there is at most a very small sex difference in general mental ability, but men clearly perform better on visuospatial tasks while women clearly perform better on tests of verbal usage and perceptual speed. In this study, we integrated these overall findings with predictions based on the Verbal–Perceptual–Rotation (VPR) model ([Johnson, W., and Bouchard, T. J. (2005a). Constructive replication of the visual–perceptual–image rotation (VPR) model in Thurstone's (1941) battery of 60 tests of mental ability. Intelligence, 33, 417–430.; Johnson, W., and Bouchard, T. J. (2005b). The structure of human intelligence: It's verbal, perceptual, and image rotation (VPR), not fluid and crystallized. Intelligence, 33. 393–416.]) of the structure of mental abilities. We examined the structure of abilities after removing the effects of general intelligence, identifying three underlying dimensions termed rotation–verbal, focus–diffusion, and memory. Substantial sex differences appeared to lie along all three dimensions, with men more likely to be positioned towards the rotation and focus poles of those dimensions, and women displaying generally greater memory. At the level of specific ability tests, there were greater sex differences in residual than full test scores, providing evidence that general intelligence serves as an all-purpose problem solving ability that masks sex differences in more specialized abilities. The residual ability factors we identified showed strong genetic influences comparable to those for full abilities, indicating that the residual abilities have some basis in brain structure and function.

Keywords: Sex differences; Residual mental abilities; Verbal and spatial abilities; General intelligence; VPR theory; Genetic and environmental influences

Sex differences in mental ability have long intrigued psychologists. Traditional analyses have examined mean and variance differences in scores on tests of general and specific abilities (Feingold, 1995). More sophisticated statistical analyses have addressed sex differences in factor structures, *g*-loadings, and differential functioning of test items that might explain the observed mean differences (Abad, Colom, Rebollo, & Escorial, 2004; Halpern, 2000; Jensen, 1998). Laboratory-based studies have made use of magnetic resonance imaging (MRI) and other techniques to investigate differences in sizes of structural brain components and physiological and neurological activity levels in various brain regions (Gur, Gunning-Dixon, Bilker, & Gur; 2002; Kimura, 1999; Sommer, Aleman, Bouma, & Kahn, 2004). The empirical performance data support the overall conclusion that there is at most a small sex difference in general ability (Colom & Lynn, 2004; Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Jensen, 1998; Lynn & Irwing, 2004), but men tend to do much better on many visuospatial tasks while women tend to do much better on many tests of verbal usage and perceptual speed (Halpern, 2004; Jensen, 1998). The empirical laboratory data provide evidence for, among other things,

---

* Corresponding author. Tel.: +1 952 473 1673; fax: +1 952 473 1998.
E-mail address: john4350@umn.edu (W. Johnson).

0160-2896/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2006.03.012

greater cerebral blood flow in women than in men (Gur & Gur, 1990), larger hypothalami in men than in women (LeVay, 1991), larger corpus collosi in women than in men (Steinmetz, Staiger, Schluag, Huang, & Jancke, 1995), greater frontal and parietal cortical complexity in women than in men (Luders et al., 2004), more bilateral activation during verbal tasks in women (Shaywitz et al., 1995) and during spatial tasks in men (Gur et al., 2000), and greater involvement of white matter in women and of gray matter in men in different brain regions (Haier, Jung, Yeo, Head, & Alkire, 2005).

Taken together, these data suggest that men and women achieve similar levels of overall intellectual processing power using different neuroanatomic and brain structural pathways, which in turn contribute to differences in more specialized abilities. There may be differences in the manner in which *g* is manifested within the sexes that parallel the differences between the sexes. This would imply that there is no single structural and functional brain system that manifests as general intelligence (*g*). It is the differences in *g*, however, that have the universal and far-reaching practical effects in predicting life outcomes. (Deary, Whiteman, Whalley, Fox, & Starr, 2004; Gottfredson, 1997). When *g* is removed from the scores on a battery of tests by statistical regression, the practical validity of the residual scores with respect to any general outcome in any general population is largely destroyed (Jensen, 1998).

This suggests that *g* is of general-purpose use, yet used to access components that vary from individual to individual. The analogy might be that every individual's *g* is associated with an intellectual toolbox, but individuals vary both in the skill with which they choose and use their tools (their *g*) and in the specific tools to which they have access. Most intellectual tasks can be accomplished using several to many different combinations of tools and different individuals will tend to use different combinations of tools, depending on their skill with the specific tools in their individual chests. But some tasks can be accomplished much better with certain tools than with others, and individual performance on these tasks depends not only on skill in tool use, but also to some degree on individual toolbox composition. Within this analogy, it appears to be common for men to rely on sets of tools that differ somewhat from those commonly relied upon by women, and vice versa. The analogy is incomplete, of course, but it makes clear the question we address in this paper, namely, what are the differences in specific tool use (mental abilities) of men and women when overall skill in tool choice and use (*g*) is removed?

Johnson and Bouchard (2005a,b) compared the fit of two prominent models of mental ability, the fluid-crystallized (Hakstian & Cattell, 1978) and verbal–perceptual (Vernon, 1964), in two large but very different batteries of tests, as well as in an important data set used to support the fluid-crystallized model (Hakstian & Cattell, 1978). The results in all three data sets clearly favored the verbal–perceptual model including a *g* factor, and the fit of this verbal–perceptual model could be substantially improved with the addition of a second-stratum (Carroll, 1993) memory factor and a third-stratum three-dimensional image rotation factor. Johnson and Bouchard (2005a) termed this enhanced model the Verbal–Perceptual–Image Rotation (VPR) model, and argued that it offers a description of the structure of mental ability that is theoretically superior to the fluid-crystallized model because it is more consistent with the idea of coordination of function across brain regions and with the known importance of brain laterality in intellectual performance. We now describe the VPR model in some detail because the overall purpose of this paper was to use it to develop and investigate sex differences in mental abilities.

The VPR model is hierarchical, with a fourth-stratum *g* factor that contributes strongly to broad third-stratum verbal, perceptual, and image rotation abilities, which in turn contribute to 8 s-stratum factors representing more specialized abilities that contribute to specific test performance (Verbal, Scholastic, Fluency, Number, Content Memory, Perceptual Speed, Spatial, and Image Rotation). In Johnson and Bouchard (2005b), which probably best articulates the model, the third-stratum verbal and perceptual abilities, though separable, were highly correlated (.80), as were the perceptual and image rotation abilities (.85). The verbal and image rotation abilities, however, were much less correlated (.41). The third-stratum Verbal factor contributed to more specialized Verbal, Scholastic, Fluency, and Numerical abilities. The third-stratum Perceptual factor contributed to more specialized Numerical, Content Memory, Perceptual Speed, and Spatial abilities. The third-stratum Image Rotation factor represented three-dimensional image rotation abilities at the second stratum as well. Both the third-stratum Verbal and Perceptual Ability factors contributed to the specialized Numerical Ability factor, with Verbal Ability making the larger contribution. The model highlights the importance of image rotation ability in the overall structure of mental abilities, and also shows its relative distinctness from the verbal abilities more closely associated with most educational systems.

In the VPR model, the contributions of the higher-order factors to the lower-order factors and specific test scores are strong, but there is clear evidence for each level of specialized abilities. There are meaningful residual mental abilities remaining after *g* is taken into account. The structure of these residual abilities and the relationships among them thus become of interest (cf., Gustafsson, 1988). One aspect of this structure seems predictable. The large contribution of *g* to the lower-stratum abilities, the observation that different individuals use different neuroanatomic and structural brain pathways to accomplish the same tasks, and the relatively low correlation between verbal and image rotation abilities imply that, absent the overall influence of *g*, residual verbal and image rotation abilities might be negatively correlated.

This is not a new idea. The existence of such a negative association and an evolutionary basis for it were suggested by Lynn (1987). This led to a debate (Ashton & Vernon, 1995; Lynn, 1990; Nyborg & Sommerlund, 1992; Vernon, 1990) about the mathematical inevitability of such a negative correlation given that *g* is defined as the variance common to a group of mental ability test scores. As noted in that debate, because each test score can be considered to be a sum of variance common to all the test scores and variance unique to it, some negative correlation among some of the variables is mathematically inevitable when the magnitude of the common variance is large relative to the components of unique variance in the variables contributing to it. In effect, *g* acts as a variable suppressing the underlying negative association between residual verbal and image rotation abilities. This does not, however, render display of the underlying relation an artifact of the approach taken to extract it. Rather, as Lynn (1990) pointed out, the pattern of the negative correlations clarifies the nature of the underlying association, thus making possible the generation of hypotheses regarding its origins and correlates. We can do this in the same way that we can gather information about the characteristics of ancient human tool users by examining the remains of their tools in archeological sites.[1] One purpose of this study was to develop a model of the structure of residual mental abilities based on the VPR model of full mental abilities, with particular focus on articulating the relation between residual verbal and image rotation abilities. Because many of the observed sex differences in mental abilities tend to involve verbal- and image rotation-related abilities (Halpern, 2000; Peters, 2005), we hypothesized that this residual association defines a dimension along which these sex differences lie.

At the same time, to the extent that our analogy is apt and two conditions exist, we can make a further prediction. The first condition is that *g*, or overall skill in tool choice and use, makes use of a problem-solving toolbox that differs from individual to individual. The second condition is that the overall usefulness of the tools and the skill in their choice and use are reasonably evenly distributed between men and women. In the presence of these conditions, *g* should tend to mask sex differences in the specialized tools that contribute to more specialized abilities. Thus, its removal from the scores on a battery of tests through statistical regression should reveal greater sex differences in the residual scores than in the original full scores, and the sex differences in the scores on the residual factor scores should also be larger than commonly observed sex differences in mental ability test scores. The second purpose of this study was to investigate these predictions. Finally, if in fact the structure of residual abilities reflects aspects of brain structure and function, we would expect residual abilities to be subject to genetic influences independent of *g*. The third purpose of this paper was to estimate the proportions of variance in the residual ability factors that can be attributed to genetic and environmental influences.

1. Method

1.1. Research participants

The 436 (188 males, 248 females) research participants for this study came from the Minnesota Study of Twins Reared Apart (MISTRA). The sample consists of adult twins who were reared apart, along with many of their spouses, partners, adoptive and biological family members, and friends. Most of the twins were separated early in life, reared in adoptive families, and reunited only in adulthood. As this is a rare occurrence, the sample is unusual. It is also unusual because it is a sample of convenience: some participants entered by contacting the study on their own initiative, while others were recruited when the investigators learned of their existence. The sample is not unusual, however, in many important ways. The participants span a broad range of characteristics in

---

[1] Ashton and Vernon raised another point in this debate. They suggested that the proper test of the existence of a negative association between verbal and image rotation abilities would be to calculate partial correlations between groups of verbal and image rotation ability tests, controlling for a *g* derived from a different body of tests. As an example of their suggested technique, they presented partial correlations between verbal and spatial parts of the WAIS, controlling for Raven scores representing *g*, with the result that the correlation between verbal and spatial abilities appeared to approach 0 rather than to be negative. The results in such a situation are heavily dependent on the degree to which the measure chosen to represent *g* loads equally on the tests chosen to represent verbal and spatial ability, data Ashton and Vernon did not present. A more thorough way to address the question would be to derive *g* from a full battery of tests and apply it to verbal and spatial factors derived from another body of tests. We could have done this and replicated it twice with the data available to us here. We do not because the results would be functionally equivalent to the analysis we do present here as the *g* factors derived from each of the three test batteries contained in our data are completely correlated (Johnson, Bouchard, Krueger, McGue, & Gottesman, 2004).

the areas of socioeconomic status, general health, personality, psychopathology, and intellectual ability.

The participants came primarily from North America, Great Britain, and Australia, though a few came from other countries. They ranged in age from 18 to 79 (mean = 42.7). Their educational backgrounds varied from less than high school to post-graduate experience, and occupations spanned a corresponding range. The sample included 128 twin pairs, 2 sets of triplets, 117 spouses of twins, and 57 other biological and adoptive family members of the twins. Initiated in 1979, the study continued to recruit reared apart pairs until 2000. In addition, some participants returned for a second assessment 7 to 12 years after the initial one. The assessment consisted of a week-long battery of tests evaluating medical and physical traits as well as psychological (cognitive abilities, personality, interests, attitudes, etc.) characteristics. Bouchard, Lykken, McGue, Segal, and Tellegen (1990) and Segal (2000) report further details on recruitment and assessment. Most of the mental ability tests were administered in blocks lasting 60 to 90 min throughout the full period of assessment. In total, there were three 3 cognitive ability batteries in the assessment that were relevant to this study. We describe them in turn.

1.2. Measures

1.2.1. Comprehensive Ability Battery (CAB)

The CAB was developed by Hakstian and Cattell (1973). It consists of 20 brief (5–6 min each) primary ability tests developed with the goal of measuring a broad range of well-replicated primary abilities. To make maximal use of available time by avoiding duplication of tasks in the extensive MISTRA assessment, 6 of the tests in the CAB were not administered to the participants (Auditory Ability, Originality, Representational Drawing, Aiming, Spontaneous Flexibility, and Ideational Fluency). In addition, we omitted the test of Esthetic Judgment from this analysis, as we deemed it at best indirectly relevant to cognitive ability. Hakstian and Cattell (1978) reported split-half and retest reliabilities from the tests ranging from .64 for Perceptual Speed and Accuracy to .96 for Memory Span. The Verbal Ability test consists of 2 completely different tasks, multiple-choice vocabulary and proverb interpretation exercises. We thus tabulated the scores on these 2 parts separately, giving us a total of 14 test scores from the CAB.

1.2.2. The Hawaii Battery, including Raven's Progressive Matrices (HB)

The HB was developed to assess familial resemblance in cognitive ability in the Hawaii Family Study of Cognition (DeFries et al., 1974; Kuse, 1977). It consists of 15 tests of primary abilities. The tests are short, requiring 3 to 10 min for administration. Again, to make maximal use of available time by avoiding duplication of tasks, 2 tests in this battery were not administered (Number Comparison and Social Perception). The battery was supplemented, however, with 4 tests from the Educational Testing Services (Cubes and Paper Folding for spatial relations, Identical Pictures for perceptual speed and accuracy, and Different Uses for fluency) in order to articulate more clearly certain factors thought likely to be important. There were thus a total of 17 tests in the battery. The original battery included a shortened printed version of the Raven (1941). In MISTRA, the Raven was presented via slides (Lykken, 1982), and administered on an untimed basis. DeFries et al. (1974) reported internal consistency and retest reliabilities for the tests ranging from .58 for Immediate Visual Memory to .96 for Vocabulary.

1.2.3. The Wechsler Adult Intelligence Scale (WAIS)

The WAIS (Weschler, 1955) is probably the best known and most widely used individually administered test of general intellectual ability. It includes Verbal and Performance subcomponents. Many of the subtests require overt verbal articulation of reasoning based on factual knowledge. Thus, for example, the WAIS Vocabulary subtest requires the examinee to articulate his/her own definitions of the words presented. There are 11 subtests of the WAIS. Internal consistency reliabilities range from .79 for Comprehension to .94 for Vocabulary (Weschler, 1955). Normed at the 1955 level, average WAIS full-scale IQ for this sample was 109.6 (range 79–140), with a standard deviation of 11.8. Using the average rates of change in WAIS scores summarized by Jensen (1998, p. 319)[2] to adjust for secular changes in IQ, the average WAIS full-scale IQ for the sample was 101.2 (range 61.1–139.9), with a standard deviation of 14.8. The adjustment increased the standard deviation because IQ was *positively* correlated with age in this sample.

Table 1 summarizes the tests administered to MISTRA participants from the three batteries.

1.3. Statistical analyses

Because we were interested in examining sex differences in means and in relationships among types of abilities, the issues of measurement and factorial

---

[2] To estimate change in full-scale IQ, we weighted the rates of change for verbal and performance IQ .6 and .4, respectively. We individually adjusted scores downward from date of assessment to 1955 and upward by age at assessment in excess of 25. Other adjustment terms yielded similar results.

Table 1
Tests included in the 3 batteries

Test	Assessment activity
Comprehensive Ability Battery	
1. Numerical Ability	Computations including fractions, decimal divisions, square roots, etc.
2. Spatial Ability	Interpretation of 2-dimensional figural rotation or reversal.
3. Memory Span	Recall of digits presented aurally.
4. Flexibility of Closure	Identification of embedded figures.
5. Mechanical Ability	Identification of mechanical principles and tools.
6. Speed of Closure	Completion of gestalt.
7. Perceptual Speed	Evaluation of symbol pairs.
8. Word Fluency	Production of anagrams.
9. Inductive Reasoning	Identification of pattern in sequences of letter sets.
10. Associative Memory	Rote memorization of meaningless pairings.
11. Meaningful Memory	Rote memorization of meaningful pairings.
12. Verbal—Vocabulary	Multiple choice among possible synonyms.
13. Verbal—Proverbs	Interpretation of proverbs.
14. Spelling	Multiple-choice identification of misspellings.
Hawaii Battery with Raven	
15. Card Rotations	Matching of rotated alternatives to probe.
16. Mental Rotation	Identification of rotated versions of 2-dimensional prepresentation of 3-dimensional objects.
17. Paper Form Board	Outline of cutting instructions to form the target figure.
18. Hidden Patterns	Identification of probe figures in more complex patterns.
19. Cubes	Identification of matched figures after rotation.
20. Paper Folding	Identification of unfolded version of a folded probe.
21. Raven	Identification of analogous figure to follow a sequence of figures.
22. Vocabulary	Multiple choice among possible meanings.
23. Subtraction/Multiplication	Completion of 2-digit subtractions and 2-digit by 1-digit multiplications.
24. Word Beginnings/Endings	Generation of words beginning and ending with specified letters.
25. Pedigrees	Identification of familial relationships within a family tree.
26. Things Categories	Generation of things that share assigned characteristics.
27. Different Uses	Generation of novel uses for specified objects.
28. Immediate Visual Memory	Recall of illustrations of common objects immediately following presentation.
29. Delayed Visual Memory	Recall of illustrations of same common objects after delay.
30. Lines and Dots	Trace of a path through a grid of dots.
31. Identical Pictures	Identification of alternative identical to probe.
Wechsler Adult Intelligence Scale	
32. Information	Recall of factual knowledge.
33. Comprehension	Explanation of practical circumstances.
34. Vocabulary	Free definition.
35. Coding	Identification of symbol–number pairings.
36. Arithmetic	Mental calculation of problems presented verbally.
37. Similarities	Explanation of likenesses between objects or concepts.
38. Digit Span	Recall of spans of digits presented aurally, both forwards and backwards.
39. Picture Completion	Identification of parts missing in pictures of common objects.
40. Block Design	Reproduction of 2-dimensional designs using 3-dimensional blocks.
41. Picture Arrangement	Chronological sequencing of pictures.
42. Object Assembly	Reassembly of cut-up figures.

invariance were germane to this study. Measurement invariance exists when a measured construct has the same measurement properties in different groups. When measurement invariance is present, we can be sure that group mean and variance differences in levels of variables marking the construct reflect actual differences in levels of the construct among the groups. To the extent that measurement invariance is not present, mean and variance differences may reflect differences in the relative importance of the marker variables used to measure the construct (Hofer, Horn, & Eber, 1997). Given the body of evidence suggesting that men and women achieve similar levels of overall intellectual processing power using different neuroanatomic and brain structural pathways, it seems likely that measurement invariance across sex is less than complete.

Measurement invariance can be assessed by using multiple group confirmatory factor analysis to investigate the degree of factorial invariance present (Meredith, 1993). We did not implement this approach here as it proved mathematically intractable due to variance differences in the scores of males and females and the modest sample size in relation to the number of tests. In general, and consistent with the relevant literature, male variance was greater than female variance for most tests, but there were some exceptions to this pattern that made the establishment of a common metric for the two groups numerically impractical. We did, however, fit separate VPR models to the 42 full age-adjusted male and female scores. The VPR model fit well in both sexes ($\chi^2 < 2^*df$, Root Mean Square Error of Approximation < .05 (RMSEA; Browne & Cudeck, 1992; see below), providing evidence for the same numbers of factors and patterns of factor loadings in the two sexes. In addition, the magnitudes of the loadings were generally similar, but there were some differences in some of the loadings that suggested that they could not readily be constrained to be equal. Perhaps not surprisingly, the greatest differences involved the Numerical Ability factor and the tests on which it loaded. Numerical Ability appeared to involve perceptual speed to a much greater degree in females than in males. We interpret the results of the analyses we describe below in light of this apparent lack of metric invariance (Horn & McArdle, 1992).

We constructed residual ability test scores based on scores adjusted for both age (regressing out age and age^2) and age and sex (regressing out age, age^2, sex, sex × age, and sex × age^2). For both, we did this by calculating the first principal component of the set of 42 adjusted tests scores to represent *g*, calculating the first principal component score for each participant, and regressing this first principal component from the adjusted test scores, retaining unstandardized residuals to preserve the variance differences among tests. This has the advantage of being straightforward and simple, but the disadvantage of introducing small bits of specific abilities into the *g* component. Where participants were missing data for particular tests, we replaced the missing data with the participants' average scores on the tests for which data were present in calculating the *g* scores to regress from their adjusted test scores. We then used maximum likelihood confirmatory factor analysis to address the first purpose of this study, which was to develop a model of the structure of residual mental abilities. We did this by estimating the VPR model in the age- and sex-adjusted residual abilities and modifying it as appropriate for the absence of the effects of general intelligence in the residual ability scores. We used the age- and sex-adjusted residual test scores for this rather than using the age-adjusted scores and considering sex as another test score in the model as recommended by Jensen (1998) because we wanted to establish a basic model of the residual abilities absent sex effects since we were unable to estimate the extent of measurement invariance directly.

In fitting the models of residual abilities, we made no explicit adjustment for the correlated nature of the observations for the twin pairs within our sample. This has little effect on parameter estimates, but it will tend to make models appear to fit better than they actually do (McGue, Wette, & Rao, 1984). To avoid inappropriate conclusions, we fit the same models using samples including only one randomly selected twin from each pair, with the identical pattern of results. The standard and most readily available chi-square measure of model fit generally indicates significant lack of fit in large samples. We thus used a chi-square statistic less than 2*degrees of freedom and RMSEA less than .05 as indications of good model fit. RMSEA measures the extent of discrepancy between the model and data per degree of freedom; thus both alternative fit statistics provide some benefit to more parsimonious models.

The second purpose of this study was to investigate the predictions that sex differences in residual ability scores should be greater than those in the full test scores. We did this by comparing the means and standard deviations for males and females in the test scores adjusted for age only, and for both age and general intelligence as represented by the first principal component. We used analysis of variance to estimate the significance of the mean differences, considering differences to be significant only at $p < .01$ due to the large number of tests made. For the differences that were significant, we estimated the effect size (standardized mean difference). In doing so, we adjusted for the fact that all of the error of measurement was contained in the residual ability scores by estimating the ratio of error of measurement in the full ability scores to the error of measurement in the residual ability scores and increasing the effect size of the mean difference by the square root of this ratio.

The third purpose of this study was to estimate the proportions of variance in the residual ability factors that could be attributed to genetic and environmental influences. We made these calculations using the residual ability scores adjusted for age only, which made it possible for us to test for sex differences in the proportions of genetic and environmental influences and to measure the correlation between genetic influences on males and females. We had little statistical power to detect any differences, however, as the numbers of twin pairs in each sex-zygosity category were small: 31 male MZ pairs, 45 female MZ pairs, 11 male DZ pairs, 26 female DZ pairs, and 15 opposite-sex pairs.

The standard quantitative genetic model for a trait is based on the assumption that the observed variance (V_p) in the trait of interest is a linear additive function of genetic (A) and shared (C) and non-shared (E) environmental variance, respectively. Symbolically, this can be expressed as,

$$V_\mathrm{p} = A + C + E.$$

In a sample of twins reared apart, the monozygotic (MZ) twins share 100% of their genetic influences, the dizygotic (DZ) twins share 50% of their genetic influences, and there are no shared environmental influences resulting from common rearing, as discussed in greater detail below. Thus, the observed covariance for MZ twins can be expressed symbolically as,

$$\mathrm{Cov}_\mathrm{MZ} = A,$$

and that for DZ twins can be expressed as,

$$\mathrm{Cov}_\mathrm{DZ} = .5{*}A.$$

Under this model, we make several other assumptions. We assume that the variance components are independent of each other, that there is no assortative mating for the trait, and that the genetic influences are additive. Genetic variance can be additive in the sense that multiple genetic influences act independently of each other. It can also be nonadditive, reflecting coordinated multiple genetic influences such as dominance and other interactive polygenic effects. Where MZ twins elicit more similar experiences than DZ's from their environments due to their greater genetic similarity, however, this is generally considered to be an expression of their genetically influenced characteristics. The independence assumption implies that there are no gene-environment interactions or correlations that would act to create differing degrees of genetic and environmental influences within different subgroups of the sample. These assumptions are generally oversimplifications of the actual situation. Violations of the assumptions do not, however, invalidate the overall approach. Rather, they render the estimates applicable only on an overall, average population-level basis, and they introduce systematic distortions in the estimates. These distortions have different effects, depending on the assumption violated. Specifically, the presence of unmeasured shared environmental influences tends to inflate estimates of genetic influence. The presence of assortative mating, known to be relatively strong for cognitive abilities (Vandenberg, 1972), for the trait tends to inflate estimates of shared environmental influences. Interaction between genetic and shared environmental influences acts to increase estimates of genetic influence; interaction between genetic and nonshared environmental influences acts to increase estimates of nonshared environmental influence. Correlation between genetic and shared environmental influences acts to increase estimates of shared environmental influence; correlation between genetic and nonshared environmental influences acts to increase estimates of genetic influence (Purcell, 2002). Thus several combinations of violations of assumptions can act to offset each other. We offer the estimates presented in this study as preliminary indications of systematic biological and genetic involvement in the residual abilities and residual ability factors, an area clearly requiring further study.

2. Results

Our first set of results concerned the model of the structure of residual abilities. Because it provided a reasonable starting point, we began by making use of the eight second-stratum factors from the VPR model (Verbal, Scholastic, Fluency, Number, Content Memory, Perceptual Speed, Spatial, and Image Rotation), though of course the VPR model's higher-order structure was irrelevant to the residual abilities because of the absence of *g*. The eight second-stratum VPR model factors provided a reasonable fit in the age- and sex-adjusted residual ability data. The Number factor was, however, clearly unnecessary, as no tests loaded significantly on it. We thus eliminated this factor and made several other adjustments, eliminating loadings on some factors for some tests and adding loadings for other tests as appropriate and relevant. In particular, the explicitly numerical tests (Arithmetic from the WAIS, Numerical Ability from the CAB, and Subtraction/Multiplication from the HB) required additional factor loadings after the elimination of the Numerical Ability factor. Arithmetic loaded on the Scholastic factor, Subtraction/Multiplication on the Fluency and Perceptual Speed factors, and Numerical Ability on the Scholastic and Perceptual Speed factors. The factor loadings from the resulting residual ability model are shown in Table 2. There were 7 second stratum factors, representing residual Verbal, Scholastic, Fluency, Content Memory, Perceptual Speed, Spatial, and Image Rotation abilities. Considering that the data were scores of residual ability after removal of the effects of general intelligence, the factor structure appeared to be very strong, though the residual variances were large, indicative of the high proportion of error variance in the residual scores. The model fit well ($\chi^2 = 1480.89$, 776 *df*, $p < .001$, RMSEA = .046).

The correlations among the second-stratum factors are shown in Table 3. They indicated clearly the hypothesized

Table 2
Standardized factor structure of mental ability battery with *g* removed (residualized abilities)

	Second-stratum factors							Third-stratum factors			
	Verbal	Scholastic	Fluency	Content Memory	Perceptual Speed	Spatial	Rotation	Verbal–Spatial	Diffusion–Focus	Content Memory	Residual
Residual test scores											
CAB Verbal — Proverbs	.50										.75
HB Vocabulary	.65										.58
Spelling	.34		.33								.76
CAB Verbal — Vocabulary	.71										.50
WAIS Vocabulary	.39	.43									.51
Information		.50									.75
Comprehension		.57									.68
Similarities		.46									.79
Arithmetic		.20									.96
Different Uses		.25	.12								.93
Things		.20	.27								.90
Word Fluency			.50								.75
Word Beginnings/ Endings			.43								.81
Digit Span			.34								.88
Memory Span			.32								.90
Meaningful Memory		.18		.42							.83
Associative Memory				.43							.82
Immediate Visual Memory				.30							.91
Delayed Visual Memory				.24							.94
Speed of Closure			.18		.24						.90
Subtraction/ Multiplication			.29		.31						.81
Numerical Ability		.26			.50						.89
Pedigrees	.18				.34						.91
Coding					.35						.88
Perceptual Speed					.58						.67
Identical Pictures					.37						.86
Lines and Dots					.18						.97
Picture Completion	.19					.25					.93
Inductive Reasoning						.14					.98
Raven						.20					.96
Picture Arrangement		.24				.30					.90
Flexibility of Closure						.19					.96
Paper Form Board					.08	.48					.78
Hidden Patterns					.21	.33					.89
Object Assembly						.45					.80
Mechanical Ability						.42					.83
Paper Folding						.48					.77
Block Design						.53					.72
Spatial Ability							.66				.57
Card Rotations							.62				.61
Cubes						.22	.33				.78

Table 2 (*continued*)

	Second-stratum factors							Third-stratum factors			
	Verbal	Scholastic	Fluency	Content Memory	Perceptual Speed	Spatial	Rotation	Verbal–Spatial	Diffusion–Focus	Content Memory	Residual
Mental Rotation						.27	.27				.79
Second-stratum factors											
Verbal								−.60	.10		
Scholastic								−.70	.50		
Fluency								−.35	−.50		
Memory								.10	−.10	1.00	
Perceptual Speed								.45	−.60		
Spatial								.70	.70		
Rotation								.60	.05		

Note: HB refers to the Hawaii Battery. We allowed residual correlations between Picture Arrangement and Picture Completion, Digit Span and Memory Span, Number and Arithmetic, Subtraction/Multiplication and Number, Immediate and Delayed Visual Memory, and Identical Pictures and Object Assembly. All factor loadings shown were significant at $p<.05$; others were fixed to 0.

negative association between residual Verbal and Image Rotation abilities. In addition, the progression of correlations down the first column of the matrix and across the bottom row clearly suggested the presence of the hypothesized verbal–image rotation dimension. The matrix was not positive definite, so it was not possible to estimate third-stratum factors directly. With heavy ridge smoothing (Joreskog & Sorbom, 2002), however, factor analysis was possible. We extracted three unrotated and thus independent factors, one representing the hypothesized rotation–verbal dimension, one representing an analogous dimension we labeled focus–diffusion of attention, and one representing content memory.[3] The solution had two residual variances on the boundary, indicating that it might not be unique. It also fit very poorly ($\chi^2=61.16$, 3 *df*, $p<.001$, RMSEA=.21), but we report it in Table 3 because we considered it important that we were able to pick up any higher-order pattern at all given the large proportion of error variance in the matrix under analysis.

Table 3
Correlations among residualized second-stratum factors, and a factor solution

	1	2	3	4	5	6	7
1. Verbal	1.00						
2. Scholastic	.47	1.00					
3. Fluency	.06	−.16	1.00				
4. Content Memory	−.12	−.24	−.25	1.00			
5. Perceptual Speed	−.42	−.81	.02	−.05	1.00		
6. Spatial	−.56	−.34	−.79	−.20	−.25	1.00	
7. Rotation	−.53	−.50	−.35	.01	.28	.43	1.00

	Standardized factor loadings		
	Rotation–Verbal	Focus–Diffusion	Content Memory
1. Verbal	−.62	−.07	.09
2. Scholastic	−.71	−.50	.09
3. Fluency	−.36	.50	−.20
4. Content Memory	.13	.21	.97
5. Perceptual Speed	.46	.61	−.13
6. Spatial	.72	−.69	−.04
7. Rotation	.62	−.01	.00

Our second set of results concerned comparison of the sex differences in the age-adjusted full and residual ability data. Table 4 shows the male and female means, standard deviations, and the effect sizes of the differences for each of the 42 mental ability test scores adjusted only to remove age effects as well as adjusted to remove both the effects of age and general intelligence. All of the differences were consistent with differences that have been observed in prior research. As hypothesized, many more mean differences were statistically significant in the residual abilities than in the full abilities, and the effect sizes of those that were significant in both were larger in the residual than in the full abilities. There was only one reversal: Lines and Dots, a test with a rather low *g*-loading (.42) but high reliability (.89), had an effect size difference favoring males of .42 for the full ability score, but only .40 for the residual ability score. Overall, the pattern of results for the residual abilities was quite dramatic: fully half the residual abilities showed no significant sex differences, but most of the differences that were significant had effect sizes in excess of .5 standard deviations, and 2

---

[3] These were not the same third-stratum factors as extracted from the full-ability VPR model (Johnson & Bouchard, 2005b).

Table 4
Sex differences in means and variances in test scores from the three batteries adjusted for age alone and for both age and *g*

Test	Age-adjusted only						Age-adjusted and *g*-adjusted					
	Male		Female		Effect		Male		Female		Effect	
	Mean	sd	Mean	sd	*p*-value	Size	Mean	sd	Mean	sd	*p*-value	Size
CAB Verbal—Proverbs	−.033	1.009	.025	.991	ns	–	−.086	.655	.066	.675	ns	–
HB Vocabulary	−.100	1.014	.077	.980	ns	–	−.164	.616	.126	.593	<.001	−.504
Spelling	−.145	1.023	.109	.966	ns	–	−.207	.715	.156	.620	<.001	−.657
CAB Verbal—Vocabulary	−.029	1.020	.023	.981	ns	–	−.084	.577	.066	.627	ns	–
WAIS Vocabulary	−.029	1.047	.022	.960	ns	–	−.073	.615	.055	.596	ns	–
Information	.163	.999	−.123	.981	.005	.287	.116	.613	−.089	.654	.002	.393
Comprehension	.041	1.043	−.031	.963	ns	–	−.002	.762	−.003	.749	ns	–
Similarities	.012	.997	−.009	1.000	ns	–	−.027	.759	.018	.729	ns	–
Arithmetic	.205	1.046	−.154	.932	<.001	.360	.163	.772*	−.122	.606*	<.001	.528
Different Uses	−.019	.946	.015	1.037	ns	–	−.077	.669*	.060	.805*	ns	–
Things	.052	1.046	−.040	.959	ns	–	.004	.784	−.003	.808	ns	–
Word Fluency	−.110	1.036	.085	.960	ns	–	−.172	.671	.133	.652	<.001	−.636
Word Beginnings/Endings	−.126	1.006	.098	.982	ns	–	−.182	.728	.142	.699	<.001	−.573
Digit Span	−.083	.961	.063	1.022	ns	–	−.111	.808	.084	.804	ns	–
Memory Span	.015	1.012	−.012	.989	ns	–	−.032	.827	.023	.799	ns	–
Meaningful Memory	−.188	.964	.144	1.001	.001	−.333	−.227	.754	.175	.759	<.001	−.597
Associative Memory	−.140	.972	.101	1.006	ns	–	−.184	.806	.133	.876	<.001	−.415
Immediate Visual Memory	−.072	.946	.055	1.033	ns	–	−.094	.905	.089	.878	ns	–
Delayed Visual Memory	−.033	1.009	.025	.991	ns	–	−.086	.655	.066	.675	ns	–
Speed of Closure	−.035	1.025	.026	.978	ns	–	−.078	.877	.058	.791	ns	–
Subtraction/Multiplication	.051	.972	−.039	1.017	ns	–	.003	.774	−.004	.805	ns	–
Numerical Ability	.132	1.077	−.101	.922	ns	–	.079	.691	−.061	.598	ns	–
Pedigrees	−.064	.927	.048	1.047	ns	–	−.137	.498*	.102	.589*	<.001	−.631
Coding	−.322	.979	.239	.944	<.001	−.563	−.352	.751	.268	.717	<.001	−.826
Perceptual Speed	−.210	.879	.155	1.052	<.001	−.366	−.256	.689	.189	.823	<.001	−.676
Identical Pictures	−.054	.990	.041	1.003	ns	–	−.100	.780	.076	.761	ns	–
Lines and Dots	.240	.945	−.179	1.000	<.001	.420	.200	.925	−.150	.871	<.001	.396
Picture Completion	.341	.925	−.256	.975	<.001	.599	.304	.767	−.226	.755	<.001	.807
Inductive Reasoning	.013	.967	−.009	1.021	ns	–	−.049	.725	.036	.711	ns	–
Raven	.001	.919	−.001	1.045	ns	–	−.019	.645	.015	.701	ns	–
Picture Arrangement	.159	.930	−.120	1.031	.006	.280	.130	.793	−.098	.907	.009	.303
Flexibility of Closure	.143	.987	−.106	.994	ns	–	.089	.743	−.064	.768	ns	–
Paper Form Board	.220	1.050	−.165	.924	<.001	.386	.166	.828*	−.125	.661*	<.001	.481
Hidden Patterns	.115	.974	−.087	1.008	ns	–	.067	.695	−.049	.709	ns	–
Object Assembly	.102	.941	−.077	1.033	ns	–	.073	.868	−.057	.892	ns	–
Mechanical Ability	.660	.918*	−.516	.714*	<.001	1.178	.636	.781*	−.495	.591*	<.001	1.431
Paper Folding	.251	.954	−.187	.990	<.001	.439	.197	.738	−.146	.746	<.001	.592
Block Design	.193	.973	−.145	.993	.001	.339	.155	.702	−.115	.701	<.001	.478
Spatial Ability	.143	.945	−.106	1.024	ns	–	.097	.798	−.072	.853	ns	–
Card Rotations	.114	1.346	−.086	.604	ns	–	.171	.541	−.071	.538	<.001	.453
Cubes	.315	.972	−.236	.952	<.001	.552	.269	.734	−.200	.762	<.001	.752
Mental Rotation	.515	.984*	−.400	.808*	<.001	.917	.481	.885*	−.372	.678*	<.001	1.039

Note: Due to the number of tests made, differences were only considered significant if $p < .01$. Effect sizes are shown only for significant differences. Effect sizes for age- and *g*-adjusted scores are adjusted for greater proportion of error variance.
*Difference in variances significant at $p < .01$.

exceeded a full standard deviation. Twelve differences favored males; 9 favored females. The average effect size favoring males was .64; that favoring females was −.61.

The factors and their loadings derived from the structural model of residual abilities based on the age- and sex-adjusted residual ability scores of course indicated no sex differences in residual abilities because we had eliminated all effects of sex from the data. Thus, the model we developed could be considered to reflect some level of brain structure and function independent of sex differences. To re-introduce and evaluate the sex effects that were

Table 5
Sex differences in residual ability factors, with comparison to sex difference in *g*

	Males		Females		Significance of difference	Effect size
	Mean	sd	Mean	sd		
Verbal	−.33	1.14	.26	1.12	<.001	−.51
Scholastic	.14	1.03	−.06	1.10	<.001	.19
Fluency	−.31	1.09	.24	1.00	<.001	−.51
Memory	−.26	.70	.21	.74	<.001	−.62
Perceptual Speed	−.25	1.05	.18	1.10	<.001	−.39
Spatial	.87	1.64	−.64	1.34	<.001	.92
Rotation	.39	.96	−.26	.96	<.001	.64
Rotation–verbal	.91	2.67	−.63	2.43	<.001	.58
Focus–diffusion	1.00	1.81	−.73	1.64	<.001	.90
Memory	−.26	.70	.21	.74	<.001	−.62
g	.08	1.00	−.06	1.00	.16	.14

present, we calculated residual ability factor scores based on the residual ability scores adjusted only for the effects of age. Table 5 shows the male and female means and standard deviations of these residual ability factor scores, along with the statistical significance levels and effect sizes of the differences in the means. Because the factor characterizing the rotation–verbal dimension was positioned so that higher rotation than verbal residual ability scores received positive values and higher verbal than rotation residual ability scores received negative values, we labeled this factor rotation–verbal. With the exception of the .19 difference in residual Scholastic ability favoring males, all of the differences in the residual abilities and their factors were moderate to large in size (Cohen, 1988). By way of comparison, the table also shows the mean difference statistics for the *g* scores. There was a difference favoring males of .14 standard deviations, but it was not statistically significant ($p=.16$).

Figs. 1, 2, 3 and 4 show the male and female distributions for *g* along with those of the 3 higher-order residual ability factor scores. The relatively even distributions of *g* across sex are apparent. The poles of the *g* and memory factor distributions reflect lower and higher levels of absolute ability. The extremes of the distributions across the rotation–verbal and focus–diffusion dimensions, however, should be interpreted differently. They do not indicate lower and higher levels of absolute ability. Rather, for the rotation–verbal dimension, negative scores indicate high residual verbal relative to residual rotation ability while positive scores indicate high residual rotation relative to verbal ability. The situation is similar for the focus–diffusion dimension: negative scores indicate high residual diffusion of attention relative to residual focus of attention while positive scores indicate high residual focus of attention relative to residual diffusion of attention.

Our third set of results consisted of estimates of genetic and environmental influences on the residual ability factors. Table 6 shows our estimates. Though our twin pairs were reared apart, the possibility exists that some similarity in their rearing homes, their association with each other subsequent to reunion, or the conditions

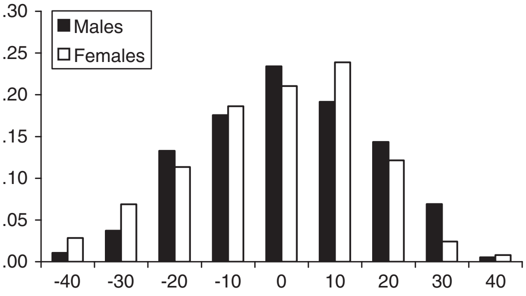

Fig. 1. Male and female distributions of *g*. Higher scores indicate higher levels of ability.

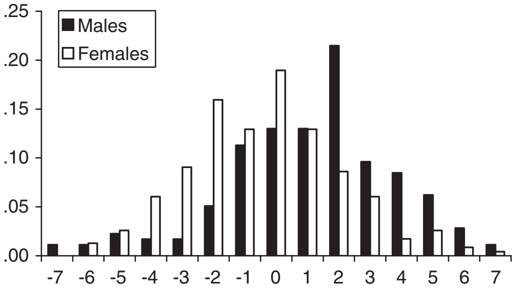

Fig. 2. Male and female distributions on rotation–verbal dimension. Negative scores indicate high residual verbal relative to residual rotation abilities. Positive scores indicate high residual rotation relative to verbal abilities.

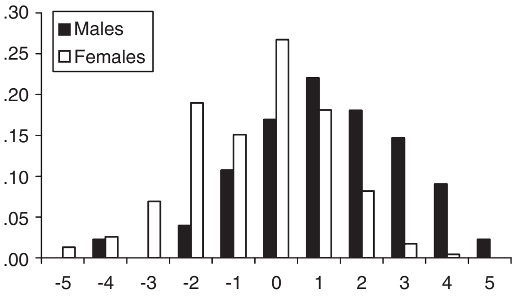

Fig. 3. Male and female distributions on focus–diffusion dimension. Negative scores indicate high residual diffusion of attention relative to residual focus of attention. Positive scores indicate high residual focus of attention relative to residual diffusion of attention.

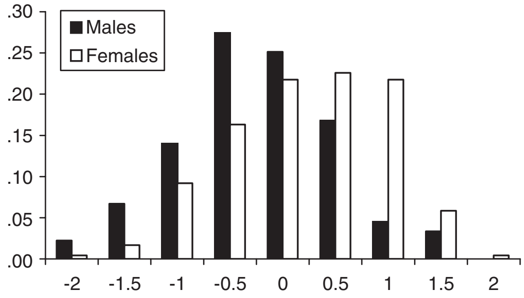

Fig. 4. Male and female distributions on memory dimension. Higher scores indicate higher levels of ability.

of their testing in our labs resulted in shared environmental influences that acted to make them similar. We thus began by fitting models considering the possibility of shared environmental influences. We were, however, able to constrain their parameters to 0 without significant loss of model fit so we discuss them no further. Despite the modest size of our twin sample and the fact that the residual ability test scores included a much higher proportion of error variance than do full ability test scores, the estimates of genetic influence on all of the residual ability factors were both substantial and highly consistent with each other. Residual content memory was the one exception, showing somewhat less genetic influence than did the other residual abilities, but this is consistent with prior research (Doetsch & Hen, 2005; Finkel, Pedersen, McGue, & McClearn, 1995; McClearn et al., 1997) and so seems unlikely to be a random consequence of error. We were unable to detect any sex differences in the magnitude of genetic influences on any of the abilities, nor did we detect any evidence that different kinds of genetic influences might be involved. We had little statistical power to detect such differences, however, due to the very modest sizes of our sex by zygosity groups.

3. Discussion

The overall purpose of this study was to use the VPR model to investigate several predictions involving sex differences in mental abilities. To do so, we focused on the structure of residual mental abilities after removing the effects of general intelligence and investigated three predictions. First, we found evidence for a hypothesized negative association between residual verbal and rotation abilities, as well as evidence that this negative relation defines a dimension along which sex differences lie. Second, we demonstrated that *g* masks sex differences where they actually lie, at the level of residual abilities.

Finally, we estimated that substantial proportions of variance in these residual abilities could be attributed to genetic influences. In fact, even though our residual ability scores included all the measurement error in the full test scores, our estimated proportions of variance attributable to genetic influences on the residual ability factors were only slightly lower than the estimates commonly reported for general intelligence in adult samples (Bouchard, 1996; Jensen, 1998). This is important because it provides evidence that the residual abilities follow systematic and regular patterns that have some biological origin in brain structure and function. The variance decomposition models we applied to make our estimates of genetic influence rely on a number of assumptions, several of which may not be realistic, making our point estimates of proportions of variance attributable to various sources of limited value. The models we applied also do nothing to articulate specific pathways by which genes may have their influence. Still, the substantial levels of genetic influence estimated indicate that, at the level of the population, across the relatively broad (by modern, Western standards) range of economic and social environments represented in our sample, the individual genome taken as a whole exerts a strong influence on adult cognitive abilities at both the general and specific levels. This suggests that, though the residual

Table 6
Estimated proportions of variance attributable to genetic and environmental influences on the residual ability factors

	Indicated proportion genetic	95% confidence interval	Indicated proportion environmental	95% confidence interval
Verbal	.60	(.44–.72)	.40	(.28–.56)
Scholastic	.64	(.49–.75)	.36	(.25–.51)
Fluency	.62	(.47–.73)	.38	(.27–.53)
Content Memory	.38	(.16–.55)	.62	(.45–.84)
Perceptual Speed	.61	(.45–.72)	.39	(.28–.55)
Spatial	.70	(.57–.79)	.30	(.21–.43)
Rotation	.59	(.43–.71)	.41	(.29–.57)
Rotation–verbal	.78	(.68–.85)	.22	(.15–.32)
Focus–diffusion	.66	(.52–.76)	.34	(.24–.48)
ContentMemory	.38	(.16–.55)	.62	(.45–.84)

abilities may not predict general life outcomes such as level of education or socioeconomic status as does *g* (Jensen, 1998), they may help predict more specific life outcomes such as selection of occupation (Gottfredson, 2002; Humphreys & Lubinski, 1996; Humphreys, Lubinski, & Yao, 1993; Shea, Lubinski, & Benbow, 2001) or development of particular skills or interests better than does *g*.

Our observations regarding the relations among the residual abilities and the sex differences in their means have to be considered in the context of our inability to establish to what degree we are measuring abilities consistently across sex and, more importantly, even across individuals within each sex. The techniques on which our findings rely are based on the assumption that measurement invariance across groups and the individuals within the groups is present. Yet it seems likely that this assumption is not met. All interpretations of our findings elide the tension that this creates. This is, however, a problem underlying all research involving mental abilities to date. In addition, performance on any task reflects learned behavior to at least some degree, and people likely differ in their prior exposure to any task with which they might be presented, no matter how novel. It is thus never possible to measure innate ability per se, and the term is highly misleading. Nevertheless, genetic differences are always reflected to varying degrees in individual test scores. Furthermore, most problems can be solved using multiple strategies. Thus it is unlikely that any specific task measures a specific ability and that ability alone.

These observations regarding specific task performance could apply as well to *g*. That is, within each individual, separately identifiable brain structures and functions must be coordinated if the individual is to behave in a coherent manner. The presence of coordination of this type within each individual could drive the consistently observed correlations among tests of mental abilities that define *g* even if the individual patterns of coordination differed from person to person. Evidence from brain scans increasingly suggests that this is the case (e.g., Haier et al., 2005; Luders et al., 2004; Toga & Thompson, 2003). This changes the focus in the use of factorial invariance (Meredith, 1993) to investigate measurement invariance from the establishment of the existence of measurement consistency to the identification of the sources of measurement inconsistency in order to identify their underlying structural and functional origins.

The VPR models we fit to the male and female full test scores in order to explore factorial invariance suggested that the biggest differences in factor loadings between the sexes involved the Number factor. This is not surprising, since sex differences in mathematical ability have been well documented (Benbow & Stanley, 1983; Geary, 1996; Hyde, Fennema, & Lamon, 1990). The idea that there may be sex differences in numerical problem-solving strategies has also been well explored (Geary, 1998; Halpern, 2000). Our finding that a Number factor, important in the VPR model as well as in most other models of the structure of abilities, is unnecessary in modeling the structure of residual abilities would appear to have two implications. First, it suggests that numerical ability is not a specialized ability in its own right, but rather that it reflects *g* and other higher-order abilities directly. There is other evidence that this is the case. For example, the numerical abilities test of the Differential Aptitude Battery predicted high school grades including English, literature, social studies, and history as well as or better than did either the verbal reasoning test or a composite of the two (Bouchard, 1978). Also, arithmetic computation was essentially interchangeable with omnibus intelligence tests in Ghiselli's (1973) review of the prediction of job training and proficiency across the entire array of occupations. (See also Lubinsky & Humphreys, 1992 and the location of Necessary Arithmetic Operations in the radix presented in Marshalek, Lohman, & Snow, 1983). Second, the apparent absence of a residual numerical ability factor, in the presence of the sex differences in factor loadings on the full Number factor, provides additional evidence supporting the existence of sex differences in numerical problem-solving strategies. Because of the great (and increasing) importance of quantitative skills in a variety of higher-level occupations, it is important to explore this issue further in future research.

Our data showed clearly the predicted negative relation between residual image rotation and verbal abilities, which is consistent with the idea that *g* acts as a general-purpose problem-solving tool suppressing varied and interesting underlying relations among the residual abilities. It would appear, for example, that a person with higher general intelligence can use that general intelligence to do something towards solving image rotation problems even when s/he has little specific image rotation ability, and such a person may even achieve a similar level of absolute success as someone with lower general intelligence who does have specific image rotation ability. The testable hypotheses implied by this are that the two people would use different strategies and parts of the brain, and their brain use patterns would more closely resemble those of people with similar levels of specific image rotation abilities than they would those with similar absolute levels of image rotation performance. In addition, we would expect that the person with little specific image rotation ability might have relatively great specific verbal ability, and vice versa for the person who does have specific image rotation ability, with similar effects on their performance on highly verbal tests. In this sense, it is reasonable to think of an image rotation–verbal dimension of specific ability underlying general intelligence, and our data provide evidence in support of this. Wittmann (2005) has compiled results very compatible with this conception using world-wide achievement tests, so the conception has practical implications as well (cf., Johnson & Bouchard, 2005b).

The existence of such a dimension implies some form of trade-off between residual image rotation and verbal abilities. What this might mean in terms of brain structure and function is not clear, but it has long been known that the two hemispheres of the brain play differential roles in performance on different intellectual tasks involving verbal and image rotation abilities (Toga & Thompson, 2003). It is also clear that there are individual differences in the degree of asymmetry between left and right brain hemispheric structures and neurochemical processes (Toga & Thompson, 2003), and it seems possible that these individual differences might be associated with the image rotation–verbal dimension. In addition, these residual abilities have shown associations with handedness, sex, hormonal levels, and the degree of brain hemispheric specialization (Halpern, 2000; Kimura, 1999; McGee, 1979; Sowell, Thompson, Holmes, Jernigan, & Toga, 1999), any and all of which could help to explain the trade-off we observed.

In addition to the predicted image rotation–verbal residual ability dimension, we observed a second residual ability dimension with similar trade-off characteristics. We had not anticipated the existence of this dimension. We termed it a focus–diffusion of attention dimension because one pole was characterized by links among residual verbal, scholastic, spatial, and image rotation abilities that could be considered to result from application of focused attention, while the other pole was characterized by links among residual fluency, content memory, and perceptual speed that might be considered to result from application of more diffuse attention to a variety of cues simultaneously. This dimension may possibly be related to the cognitive processes that the neuropsychologist Luria (1966, 1973, 1980) proposed underlie mental task performance. He suggested that, working in concert, the brain stem and reticular activating system provide the brain with the appropriate level of arousal for focused attention and resistance to distraction, the occipital–parietal and frontal–temporal areas of the brain receive, analyze, and store incoming sensory information, and the frontal lobes of the brain program and regulate behavior. This proposition was used to develop the Planning, Attention, Simultaneous, and Successive theory of cognitive abilities (Das, Naglieri, & Kirby, 1994), and the Cognitive Assessment System (CAS; Naglieri & Das, 1997) to assess these abilities. The CAS has shown substantial relations to academic achievement (Naglieri, 1999), informative profiles for diagnosing attention-deficit-hyperactivity disorder and other learning disabilities in children (Naglieri, 1999), and sex differences favoring females in the Planning and Attention scales (Naglieri & Rojahn, 2001) that require abilities similar to those we have characterized as requiring diffusion of attention. As written, the CAS assesses full rather than residual abilities, so the question of a trade-off between diffusion and focus of attention does not naturally arise. It would be interesting, however, to explore the possibility of such a trade-off in future research using this instrument. Interestingly, two scales included in the CAS, Simultaneous and Successive Processing, have parallels in the residual ability factors we identified as well: the Simultaneous scale includes two of three tests that combine verbal and rotation abilities, and the Successive scale consists essentially of memory tests.

It is important to emphasize that the data that revealed the trade-off dimensions of residual image rotation–verbal abilities and focus–diffusion of attention were adjusted to remove all effects of sex. Thus, the dimensions themselves are not sources of sex differences, and we should expect to find both men and women lying at all points along these dimensions. Still, the dimensions do appear to articulate sex differences in cognitive abilities as well. It seems possible that hormones could act to influence placement of individuals along these dimensions by nudging individual placement in the direction more common to the individual's sex. Thus, female hormones might nudge women's residual abilities toward the verbal and diffusion of attention poles, and male hormones might nudge men's residual abilities toward the image rotation and focus of attention poles, all else being equal. This would be consistent with the observation that female-to-male transsexuals show improved image rotation ability and male-to-female transsexuals perhaps show increased verbal facility (Slabbekoorn, van Goozen, Megens, Gooren, & Cohen-Kettenis, 1999; van Goozen, Cohen-Kettenis, Gooren, Frijda, & Vandepoll, 1995).

In conclusion, in this study we have presented evidence supporting the idealized notion of general intelligence as a general-purpose mechanism that accesses a toolbox made up of components that vary from individual to individual. Though everyone clearly has most if not all of the same tools, individuals appear to differ not only in the skill with which they use their tools, but also in the specific tools they habitually use. For some of the more specific tools, it would appear that using one tool means failing to use another. To the extent that this analogy is apt, it has potential implications for explaining often mixed findings from laboratory-based studies using magnetic resonance imaging and other techniques to examine brain structural, physiological, and neurological functioning because it suggests that there are individual differences in the processes used in performance of many tasks that are related to individual differences in underlying structural and functional capabilities. The analogy also has practical implications for education and career selection. Performance on image rotation tasks is known to predict success in fields such as airplane piloting, engineering, physical sciences, and fine arts better than does general intelligence, and especially verbal ability (Gottfredson, 2002; Humphreys & Lubinski, 1996; Shea et al., 2001). What has perhaps not been recognized is that inclusion of verbal ability in assessments used to recruit individuals to those fields may actually act to impair efforts to select those with the talents most relevant to the jobs in question.

Acknowledgements

We thank Nancy L. Segal, Margaret Keyes, and Kathy Corson for their many hours spent interviewing and administering tests to the participants of MISTRA. We also thank the numerous co-investigators, MISTRA staff over the years, and the MISTRA participants themselves for making the project possible and for their enthusiastic participation. Wendy Johnson was supported by a University of Minnesota doctoral dissertation fellowship during preparation of this manuscript.

References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences*, *36*, 1459−1470.

Ashton, M. C., & Vernon, P. A. (1995). Verbal and spatial abilities are uncorrelated when *g* is controlled. *Personality and Individual Differences*, *19*, 399−401.

Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science*, *222*, 1029−1031.

Bouchard, T. J. (1978). Review of the Differential Aptitude Test. In O. K. Buros (Ed.), *The eight measurements yearbook*, *Vol. 1.* (pp. 655-658). Highland Park, NJ: The Gryphon Press.

Bouchard, T. J. (1996). Behaviour genetic studies of intelligence, yesterday and today: The long journey from plausibility to proof. *Journal of Biosocial Science*, *28*, 527−555.

Bouchard, T. J., Lykken, D. T., McGue, M., Segal, N. L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota Study of Twins Reared Apart. *Science*, *250*, 223−228.

Browne, M., & Cudeck, R. (1992). Alternative methods of assessing model fit. *Sociological Methods and Research*, *21*, 230−258.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis*, 2nd ed. Hillsdale, NJ: Erlbaum.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence in 12–18-year-olds. *Personality and Individual Differences*, *36*, 75−82.

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes.* Needham Heights, MA: Allyn and Bacon.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. C., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey of 1932. *Intelligence*, *31*, 533−542.

Deary, I. J., Whiteman, M. C., Whalley, L. J., Fox, H. C., & Starr, J. M. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 130−147.

DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. G., et al. (1974). Near identity of cognitive structure in two ethnic groups. *Science*, *183*, 338−339.

Doetsch, F., & Hen, R. (2005). Young and excitable: The function of new neurons in the adult mammalian brain. *Current Opinion in Neurobiology*, *15*, 121−128.

Feingold, A. (1995). The additive effects of differences in central tendency and variability are important in comparisons between groups. *American Psychologist*, *50*, 5−13.

Finkel, D., Pedersen, N. L., McGue, M., & McClearn, G. E. (1995). Heritability of cognitive abilities in adult twins: Comparison of Minnesota and Swedish data. *Behavior Genetics*, *25*, 421−431.

Geary, D. C. (1996). Biology, culture, and cross-national differences in mathematical ability. In R. J. Sternberg, & T. Ben-Zeev (Eds.), *The nature of mathematical thinking* (pp. 145−172). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Geary, D. C. (1998). *Male, female.* Washington, D.C.: American Psychological Association.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, *26*, 461−477.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79−132.

Gottfredson, L. S. (2002). The challenge and promise of cognitive career assessment. *Journal of Career Assessment*, *11*, 115−135.

Gur, R. C., Alsop, D., Glahn, D., Petty, R., Swanson, C. L., Maldjian, B. I., et al. (2000). An fMRI study of sex differences in regional activation to a verbal and a spatial task. *Brain and Language*, *74*, 157−170.

Gur, R. C., Gunning-Dixon, F., Bilker, W. B., & Gur, R. E. (2002). Sex differences in temporo–limbic and frontal brain volumes of healthy adults. *Cerebral Cortex*, *12*, 998−1003.

Gur, R. E., & Gur, R. C. (1990). Gender differences in cerebral blood flow. *Schizophrenia Bulletin*, *16*, 247−254.

Gustafsson, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 35−71). Hillsdale, NJ: Erlbaum.

Haier, R. J., Jung, R. E., Yeo, R. A., Head, K., & Alkire, M. T. (2005). The neuroanatomy of general intelligence: Sex matters. *NeuroImage*, *25*, 320−327.

Hakstian, A. R., & Cattell, R. B. (1978). Higher stratum ability structures on a basis of twenty primary mental abilities. *Journal of Educational Psychology*, *70*, 657−669.

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (Third ed.). Mahwah, New Jersey: Lawrence Erlbaum Publishers.

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science*, *13*, 135−139.

Hofer, S. M., Horn, J. L., & Eber, H. W. (1997). A robust five-factor structure of the 16PF: Strong evidence from independent rotation and confirmatory factorial invariance procedures. *Personality and Individual Differences*, *23*, 247−269.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117−144.

Humphreys, L. G., & Lubinski, D. (1996). Assessing spatial viasualization: An underappreciated ability for many school and work settings. In C. P. Benbow, & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 116−140). Baltimore, MD: John Hopkins University Press.

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial ability in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology*, *78*, 250−261.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2), 139−155.

Jensen, A. (1998). *The g Factor.* Westport, CN: Praeger.

Johnson, W., & Bouchard, T. J. (2005). Constructive replication of the Visual–Perceptual–Image Rotation (VPR) Model in Thurstone's (1941) Battery of 60 Tests of Mental Ability. *Intelligence*, *33*, 417−430.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It's verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393−416.

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence*, *32*, 95−107.

Joreskog, K., & Sorbom, D. (2002). *LISREL 8.53: User's reference guide.* Chicago: Scientific Software International.

Kimura, D. (1999). *Sex and cognition.* Cambridge, MA: MIT Press.

Kuse, A. R. (1977). *Familial resemblances for cognitive abilities estimated from two test batteries in Hawaii.* University of Colorado at Boulder.

LeVay, S. (1991). A difference in hypothalamic structure between heterosexual and homosexual men. *Science*, *253*, 1034−1037.

Lubinski, D., & Humphreys, L. G. (1992). Some bodily and medical correlates of mathematical giftedness and commensurate levels of socioeconomic status. *Intelligence*, *16*, 99−115.

Luders, E., Narr, K. I., Thompson, P. M., Rex, D. E., Jancke, L., Steinmetz, H., et al. (2004). Gender differences in cortical complexity. *Nature Neuroscience*, *7*, 799−800.

Luria, A. R. (1966). *Higher cortical functions in man.* New York: Basic Books.

Luria, A. R. (1973). *The working brain: An introduction to neuropsychology.* New York: Basic Books.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York: Basic Books.

Lykken, D. T. (1982). Research with twins: The concept of emergenesis. *Psychophysiology*, *19*(4), 361−373.

Lynn, R. (1987). The intelligence of the mongoloids: A psychometric, evolutionary, and neurological theory. *Personality and Individual Differences*, *8*, 813−844.

Lynn, R. (1990). Negative correlations between verbal and visuospatial abilities: Statistical artifact or empirical relationship? *Personality and Individual Differences*, *11*, 755−756.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*, 481−498.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107−127.

McClearn, G. E., Johansen, B., Berg, S., Pedersen, N. L., Ahern, F., & Petrill, S. A. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*, *276*, 1580−1585.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influence. *Psychological Bulletin*, *86*, 889−918.

McGue, M., Wette, R., & Rao, D. C. (1984). Evaluation of path analysis through computer simulation: Effect of incorrectly assuming independent distribution of familial correlations. *Genetic Epidemiology*, *1*, 255−269.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 525−543.

Naglieri, J. A. (1999). How valid is the PASS theory and CAS? *School Psychology*, *28*, 145−162.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system.* Itasca, IL: Riverside.

Naglieri, J. A., & Rojahn, J. (2001). Gender differences in planning, attention, simultaneous, and successive (PASS) cognitive processes and achievement. *Journal of Educational Psychology*, *93*, 430−437.

Nyborg, H., & Sommerlund, B. (1992). Spearman's *g*, the verbal performance balance, and brain processes: The Lynn–Vernon debate. *Personality and Individual Differences*, *13*, 1253−1255.

Peters, M. (2005). Sex differences in the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and cognition*, *57*, 176−184.

Purcell, S. (2002). Variance component models for gene-environment interaction in twin analysis. *Twin Research*, *5*(6), 554−571.

Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology*, *19*, 137−150.

Segal, N. L. (2000). *Entwined lives: Twins and what they tell us about human behavior.* New York: Plume.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Constable, R. T., Skudlarski, P., Fullbright, R. K., et al. (1995). Sex differences in the functional organization of the brain for language. *Nature*, *373*, 607−609.

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604−614.

Slabbekoorn, D., van Goozen, S. H. M., Megens, J., Gooren, L. J. G., & Cohen-Kettenis, P. T. (1999). Activating effects of cross-sex hormones on cognitive functioning: A study of short-term effects in transsexuals. *Psychoneuroendocrinology*, *24*, 423−447.

Sommer, I. E. C., Aleman, A., Bouma, A., & Kahn, R. S. (2004). Do women really have more bilateral language representation than men? A meta-analysis of functional imaging studies. *Brain*, *127*, 1845−1852.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, *2*, 859−861.

Steinmetz, H., Staiger, J. F., Schluag, G., Huang, Y., & Jancke, L. (1995). Corpus callosum and brain volume in women and men. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*, 1002−1004.

Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews–Neuroscience*, *4*, 37−48.

van Goozen, S. H. M., Cohen-Kettenis, P. T., Gooren, L. J. G., Frijda, N. H., & Vandepoll, N. E. (1995). Gender differences in behavior Activating effects of cross-sex hormones. *Psychoneuroendocrinology*, *20*, 343−363.

Vandenberg, S. G. (1972). Assortative mating, or who marries whom. *Behavior Genetics*, *2*, 127−157.

Vernon, P. (1964). *The Structure of human abilities.* London: Muthen and Co., Ltd.

Vernon, P. A. (1990). The effect of holding *g* constant on the correlation between verbal and nonverbal abilities: A comment on Lynn's "The intelligence of the mongoloids…" (1987). *Personality and Individual Differences*, *11*, 751−754.

Weschler, D. (1955). *Manual for the Weschler Adult Intelligence Scale.* New York: The Psychology Corporation.

Wittmann, W. W. (2005). Group differences in intelligence and related measures. In O. Wilhelm, & R. Engle (Eds.), *Understanding and measuring intelligence* (pp. 223−239). London: Sage.