Journal of Biosocial Science

http://journals.cambridge.org/JBS

Sex differences and IQ

N. J. Mackintosh

Journal of Biosocial Science / Volume 28 / Special Issue 04 / October 1996, pp 558 - 571
DOI: 10.1017/S0021932000022586, Published online: 31 July 2008

Link to this article: http://journals.cambridge.org/abstract_S0021932000022586

How to cite this article:
N. J. Mackintosh (1996). Sex differences and IQ. Journal of Biosocial Science, 28, pp 558-571 doi:10.1017/S0021932000022586

Session 4

SOCIAL AND GROUP DIFFERENCES

Chairman: W. Milo Keynes

J. biosoc. Sci. (1996) **28**, 559–571

SEX DIFFERENCES AND IQ

N. J. MACKINTOSH

Department of Experimental Psychology, University of Cambridge

Introduction

Of all the controversial issues surrounding IQ tests, few have generated more heat and less light than the question whether different groups in our society differ in average IQ. Those who study, and find evidence of, such differences believe that they are addressing a legitimate scientific issue and one whose implications, however unpalatable they may be, society ignores at its peril. Their critics insist that the question has no intrinsic scientific interest whatsoever, and conclude that those who study it are either naive, deceiving themselves or only too happy to find justification for their own comfortable position within an unjust and unequal society. In the case of sex differences in IQ scores, at least, one could argue that both sides have got it largely wrong. With one exception, there is relatively little social or political implication in such differences as there are. More importantly, however, arguments will be presented against the critics' view that such research is of no intrinsic scientific interest. On the contrary, it has already helped to answer some important scientific questions about the nature of IQ, and has the potential to answer more. Unfortunately, those who engage in such research have usually failed to see this.

First, however, the question whether there are sex differences in average IQ must be considered. This is rather less readily answered than one might have supposed, at least in part because poor scholarship has fostered a historical myth.

Equality of the sexes: fact or artefact?

'The one exception to the general rule that different groups or populations usually differ in average IQ is that both sexes have approximately the same average IQ on most tests. This is not, however, a true empirical finding but a consequence of the manner in which the tests were first constructed . . . the two sexes were *defined* to have equal intelligence rather than *discovered* to have equal intelligence' (Evans & Waite, 1981, p. 168).

This is a popular argument, endorsed by numerous commentators (e.g. Garcia, 1981; Rose, Kamin & Lewontin, 1984). But although it contains a small grain of truth, it is a serious misrepresentation of the history of IQ testing.

Like many men in his generation, Francis Galton had little doubt that men were more intelligent than women. But, to their credit, neither Cyril Burt nor Lewis Terman shared this prejudice. Both believed, moreover, that the question of any possible sex differences in intelligence was amenable to straightforward, empirical enquiry. Burt & Moore (1912) devised a wide variety of tests for measuring perceptual, motor, associative and reasoning processes in various samples of schoolchildren of both sexes, and correlated the scores they obtained from their tests with assessments of the children's 'general intelligence' provided by their teachers. They found a variety of differences in their test results, some haphazardly favouring boys, others favouring girls, but most very much smaller than the differences measurable in various physical characteristics. The one consistent trend was a significant negative correlation between the size of the sex difference on a test and that test's correlation with teachers' assessments of general intelligence, leading Burt & Moore to conclude that 'The higher the process and the more complex the capacity, the smaller, on the whole, become the sex-differences' (Burt & Moore, 1912, p. 379).

That there was essentially no difference between the sexes in general intelligence was confirmed by Terman. Contrary to the implication of Evans & Waite's remarks (1981) it is clear that neither Terman, nor Binet before him, had given any thought at all to the question of possible sex differences in deciding what items to include or exclude from their tests. The result was, as Terman wrote in his introduction to the Stanford-Binet test, that he could use the scores of his standardisation sample of approximately 1000 boys and girls, aged 4–16, to provide an empirical answer to the question.

> 'Many hundreds of articles and books of popular or quasi-scientific nature have been written on one aspect or another of this question of sex difference in intelligence; but all such theoretical discussions taken together are worth less than the results of one good experiment. Let us see what our 1000 IQs have to offer towards a solution of the problem . . . When the IQs of the boys and girls were treated separately there was found a small but fairly constant superiority of the girls up to the age of 13 years, at 14 however the curve for the girls dropped below that for boys . . . however the superiority of girls over boys is so slight . . .that for practical purposes it would seem negligible' (Terman, 1916, pp. 69–70).

Subsequently, the question of sex differences did receive some consideration during construction of new tests; in their introduction to the first revision of the Stanford-Binet, Terman & Merrill wrote: 'a few tests in the trial batteries which yielded largest sex differences were early eliminated as probably unfair' (1937, p. 34); and Wechsler comenting on Terman & Merrill's procedure merely reported 'we have done the same' (1939, p. 106). But the effect of this was not to abolish all sex differences in performance, and in the Wechsler-Bellevue test (the forerunner of the WAIS) as in the original Stanford-Binet, there remained a small difference in overall scores in favour of women, leading Wechsler to say 'We have more than a "sneaking suspicion" that the female of the species is not only more deadly but also more intelligent than the male' (Wechsler, 1939, pp. 106–7).

It should be clear, therefore, that IQ tests were not designed from the outset to yield equal scores for the two sexes. Terman thought he was making an empirical discovery in 1916 and, whatever the generality or validity of his conclusion, he was surely right. It seems probable that his original finding that there was essentially no difference between the two sexes had some influence on the subsequent practice of test construction, justifying the rejection of occasional items or sub-tests that seemed to run counter to the general rule. The revised version of the Stanford-Binet, on the other hand, found a small but consistent difference in favour of males—which Terman & Merrill attributed to problems with the standardisation sample. But unless he and Wechsler were carefully concealing what they were up to, their initial tests discovered that there were no more than trivial differences in overall IQ between the sexes, and that such differences as there were favoured women.

Having set the historical record straight, it might seem that the obvious next step will be to ask whether Terman's and Wechsler's original conclusions have been confirmed by subsequent research. But that too is easier said than done. For the small grain of truth in the historical myth is that both Terman and Wechsler also found that, although the overall difference between the sexes was trivial, there were some items or sub-tests on which females consistently did better than males, and others on which males consistently obtained higher scores than females. The two happened more or less to cancel each other out to yield approximate overall equality. But it now seems obvious that a judicious choice of sub-tests for inclusion in one's test battery could yield any outcome one wanted. Indeed, feminists inclined to see male conspiracies lurking everywhere may think that this is just what has happened with the revisions of the Wechsler tests, for both the WISC-R and WAIS-R now yield a significant overall male superiority (Jensen & Reynolds, 1983; Reynolds *et al.*, 1987).

The differences in the original standardisation samples for the two tests were small, 1·7 points for the WISC-R and 2·2 for the WAIS-R. But Lynn (1994) reviewed a number of other large scale studies of the Wechsler tests which consistently found a significant male superiority, averaging 2·35 points on the WISC and 3·08 points on the WAIS. There can be little doubt that the sex difference on these tests is reliable—and slightly larger for adults on the WAIS than for children on the WISC. But is it real, or at least, typical? One might suppose that the best way to answer the question is to look at the results obtained with other general IQ test batteries on large, representative samples of the population. The answer turns out to be somewhat equivocal. Thus Herrnstein & Murray (1994) obtained the test scores of some 12,000 teenagers and young adults on the AFQT test and found a difference of 0·9 IQ points in favour of men. But Lubinski & Humphreys (1990) analysed the test scores of some 100,000 16-year-old American schoolchildren, and found a difference of 0·3 IQ points in favour of girls; while the 1980 standardisation of the Differential Aptitude Tests, on a representative sample of American 14–18-year-olds, yielded an overall difference of 0·8 IQ points in favour of females (Feingold, 1988).

This suggests the possibility that the differences observed in the Wechsler tests are indeed a consequence of the particular make-up of those tests. As Lynn (1994) notes, the WAIS and WISC consist of a relatively arbitrary collection of different sub-tests, and to suppose that the average of this arbitrary set of scores represents 'general intelligence' is not necessarily justified. He is surely right. But he does not appear to take his argument to its appropriate logical conclusion. Instead, he suggests that the Wechsler tests actually underestimate the true difference between the sexes—because they do not include the type of spatial and mechanical reasoning test at which males particularly excel. While this is undoubtedly true, Lynn's conclusion does not necessarily follow: why might one not equally argue that the results could be biased towards female superiority by including more items at which they excel? And why have other test batteries yielded quite different results?

Table 1. Sex differences on five tests of the Differential Aptitude Test 1980 Standardisation (from Feingold, 1988)

	Language	Spelling	Clerical speed and accuracy	Mechanical reasoning	Space relations
Effect size*	−0·45	−0·40	−0·34	+0·76	+0·15

*Effect size in *d* (standard deviation units). A positive sign indicates male scores higher than female, a negative sign female scores higher than males.

It looks as if the question whether the two sexes differ in overall IQ is uninteresting, because if by overall IQ is meant the average score obtained on a large diverse test battery, the answer given will depend on the test battery used. But this is not the same as saying that the approximate equality of the sexes in overall IQ is an artefactual consequence of a decision to balance items favouring one sex with items favouring the other. On the contrary, the import is to question whether the concept of 'overall IQ' is a scientifically useful one.

Overall IQ is simply the average of scores on a heterogeneous set of tests. And that they are heterogeneous is surely proved by the fact that on some there seem to be reliable sex differences in one direction, on others a difference in the opposite direction, and on yet others no difference at all. Both Terman and Wechsler found that males tended to out-score females on tests of mental arithmetic and spatial reasoning, while females out-scored males on some verbal tests and on measures of perceptual and clerical speed. These differences have often been confirmed with other tests of these sorts of abilities. One clear example of this is provided by the Differential Aptitude Test (DAT). The sex differences found in the 1980 American standardisation sample for five of these tests are shown in Table 1. Females out-scored males on the two verbal tests and on the clerical speed test, while males out-scored females on the two spatial tests. The differences, which range from about 3 to over 13 IQ points, are not trivial, and are relatively uniform across the five age groups of the sample—although it is true that the difference on the two spatial tests was 2 or 3 points larger in 18-year-olds than in 14-year-olds.

Hedges & Nowell (1995) have analysed the results of five large scale American surveys conducted over the past 10–35 years (they include the two data bases used by Herrnstein & Murray (1994) and Lubinski & Humphreys (1990) whose overall test results were mentioned above). A summary of their results for seven different types of tests is shown in Table 2. With the exception of mechanical reasoning, the differences are rather smaller than those from the DAT shown in Table 1. They are nevertheless virtually all significant.

The DAT data seem to overestimate the magnitude of female superiority on verbal tests. According to Hyde & Linn's (1988) meta-analysis, the average sex difference on verbal tests in studies conducted since 1973 was only 0·10 *d* (1·5 IQ points) in favour of females. And females actually have a significantly lower average verbal IQ score than

Table 2. Sex differences on seven types of ability test (from Hedges & Nowell, 1995)

Test	Reading comprehension	Vocabulary	Mathematics	Perceptual speed	Associative memory	Spatial ability	Mechanical reasoning
Median effect size*	−0·09	+0·02	+0·17	−0·23	−0·26	+0·19	+0·77

*Effect size in *d* (standard deviation units), with a positive sign indicating male superiority, and a negative sign female superiority. The results shown are the median of the studies surveyed by Hedges & Nowell (for different types of test, the number of studies ranged from two to five).

males on the Wechsler tests (by 2·2 points in the WAIS-R standardisation sample). Although Lynn suggests that 'we can do no better than use the verbal IQs of the standardisation samples of the WAIS and WAIS-R' (1994, p. 260) as an estimate of verbal ability, the significant male superiority here remains surprising. It does not accord with the conclusions of Hyde & Linn's (1988) meta-analysis, nor with the results of Hedges & Nowell's (1995) survey. (A possible explanation is suggested by an informal observation made by the author in giving the WAIS information sub-test to Cambridge undergraduates for the past 10 years as an illustrative exercise: males obtain marginally higher scores than females, but this small average difference conceals some very much larger differences in individual items. Here is a case where a judicious choice of items to include could produce any result one wanted.)

The DAT also contains a numerical ability test on which, in 1980, females out-scored males by 1·5 points on an IQ scale—in sharp contrast to the results reported by Hedges & Nowell (1995) for a wider range of mathematics tests. However, Hyde, Fennema & Lamon's (1990) meta-analysis of mathematical performance also suggested that in recent studies of unselected samples of students, females obtained slightly higher average scores than males. Whatever the true picture on verbal and mathematical tests, it is certain that the DAT space relations test and Hedges & Nowell's summary underestimate the magnitude of male superiority on at least some types of spatial test. Tests of three-dimensional mental rotation reliably yield a sex difference of some 13–14 IQ points (Masters & Sanders, 1993).

There is, however, a second even more important conclusion suggested by the literature: the sexes do not differ on other types of IQ test designed to measure a more general reasoning ability. The DAT has two such tests, abstract and verbal reasoning. Although Lynn (1994) has reported data from other countries' standardisations, suggesting that males out-score females, the difference over all age groups in the most recent American standardisation sample was 0·3 points on the verbal, and 0·6 points on the abstract reasoning test—both in favour of women; in the oldest age group (18-year-olds) there was no difference at all on verbal reasoning and one of 0·3 points in favour of males on abstract reasoning (Feingold, 1988). Hedges & Nowell (1995) summarised the results of two of their surveys that gave a non-verbal reasoning test: one found a 0·6-point advantage for males, the other a 3·3-point advantage for females. The paradigm test of non-verbal, abstract reasoning ability is, of course, Raven's Matrices. Large scale studies of Raven's tests have yielded all possible outcomes, male superiority, female superiority and no difference (Court, 1983). In the most recent British and Irish standardisations of the tests, however, with children aged between 5 and 16, the only difference found at any age was one of significant female superiority at age 11 in the British data (Raven, 1981). In the two oldest age groups, the sex difference was essentially zero. The Child Health and Education Study, with a nationally representative sample of 14,000 10-year-old children, found a similar small difference in favour of girls on the Matrices Test of the British Ability Scales (Mackintosh & Mascie-Taylor, 1986). And Conrad (1979) found no sex difference in performance on Raven's Matrices in a sample of 450 15–16-year-old deaf children. Finally, Flynn (personal communication) reports that there is no sex difference in performance on Raven's Matrices in Israeli 18-year-olds.

So, do the sexes differ in IQ? It depends on the IQ test. Do they differ in intelligence? It depends on what we mean by intelligence. For Lynn (1994), the answer appears to be quite simple. Although, as noted, he rejects the idea that intelligence might be the average of the scores obtained on the various sub-tests of the WAIS or WISC, he is quite happy to accept that it is the arithmetic average of the scores obtained on a rather wider range of tests. But what tests? He believes that theoretical justification can be found for including tests of verbal, spatial and reasoning abilities. He is surely right in supposing that these are three important dimensions of human abilities. But one can certainly question the measures of these that Lynn chooses to employ. He relies on the Wechsler verbal scale as a measure of verbal ability, the verbal and abstract reasoning scores from all standardisations of the DAT since 1947 as a measure of reasoning ability, and the average effect size on three different types of spatial tests in Linn & Peterson's (1985) meta-analysis as a measure of differences in spatial ability. The average of these three gives him an overall IQ difference of some 4 points in favour of males. But as already seen, the Wechsler tests are unusual in their estimate of male superiority on tests of verbal ability. Moreover, most sex differences on the DAT have decreased since 1947; the male superiority on verbal and abstract reasoning in earlier standardisations has now, as also seen, given way to marginal female superiority; there can be little justification for using older data, when more recent data are available. If the results of Hedges & Nowell's much larger, relatively recent survey (Table 2) were used, a quite different picture would emerge: the average of the sex differences on tests of reading comprehension and vocabulary (verbal ability), non-verbal reasoning, and spatial ability gives only 0·2 point advantage for males.

But why are verbal, spatial and reasoning tests the only ones that should be averaged to give an overall IQ score? Why not include measures of perceptual speed which, in Hedges & Nowell's data, would tilt the balance in favour of females? Why not use the arithmetic average of all eight tests from the DAT which yield a female advantage of 0·8 IQ points? The answer, it is clear from Lynn's (1994) paper, is that he believes that males must have a substantially higher average IQ than females because they have a bigger brain. It is true that males have larger brains than females of the same body size (Ankney, 1992). One can also accept that there is a moderate correlation between brain size and IQ (Rushton & Ankney, 1995). But whether these facts have any implication for the relative levels of intelligence of men and women seems distinctly questionable. The size of the brain is correlated with the size of the body, both in humans and other animals. Since larger animals (e.g. elephants) are not necessarily thought to be more intelligent than smaller ones (e.g. humans), any comparison between groups that differ in average body size, such as elephants and people or men and women, must be a comparison of brain-size:body-weight ratios. But such comparisons are necessarily fraught with difficulties of interpretation, since any difference in the ratio may be a reflection of differences in constraints on the size of the body rather than a difference in the size of the brain. The point is best illustrated by a rather distant comparison. Birds, as a group, have a higher brain-size:body-weight ratio than fish, and on average a bird weighing the same as a fish will have a significantly larger brain. Although some comparative psychologists have inferred that birds must therefore be more intelligent than fish, it is easy to see that the inference is invalid. Fish, living in a relatively dense medium like water, have fewer constraints on their body size than birds, who must support their weight in the air. Hence the difference in ratio of brain to body may reflect a difference in bodies, not in brains. The same argument must apply to a comparison of the sexes in a sexually dimorphic species such as *Homo sapiens*.

What is even more questionable, however, and for the present purpose also a great deal more important, is whether the average of any heterogeneous set of test scores is as informative as the original scores themselves. The fact is that there are relatively reliable sex differences on many types of IQ test, some favouring males, other females, while on other types of test there are no differences at all. It seems more sensible to stick reasonably closely to these moderately well established facts. They certainly have implications for any attempt to specify what it is that IQ tests measure. Clearly, they cannot all be measuring the same thing. The pattern of sex differences observed reinforces the distinction between, at the least, such broad factors as verbal, spatial, abstract reasoning and perceptual speed. Indeed, it goes further than this, for verbal and spatial tests themselves yield a wide range of sex differences, implying that different tests within these broad domains must be tapping distinct abilities and skills.

But, it will be argued, performance on any one of these tests is correlated with performance on any other: hence, general intelligence, defined as *g*, or the first principal component of any suitably diverse battery of mental tests. Is there not a sex difference in *g*? The answer is still: it depends. The general factor extracted from the Wechsler tests yields a difference between the sexes, in favour of males, that is actually some 40% larger than their difference in average scores (Jensen & Reynolds, 1983). Thus the 2–3-point advantage for males in overall IQ on the WAIS would again, to Lynn's satisfaction, translate into a 4-point male advantage on the general factor extracted from the test. The reason for this is that the various sub-tests of the WAIS differ rather widely in their loadings on this general factor and there is some tendency for males to do better on sub-tests with higher *g*-loadings. The effect of this is to give more weight to those sub-tests that yield greater male superiority.

There can be no possible doubt, however, that other test batteries would yield quite different results. For example, the difference in overall scores reported by Herrnstein & Murray (1994) for the AFQT test is only 0·9 IQ points; males have higher scores on two of the sub-tests comprising this test battery, females on the other two (see Hedges & Nowell, 1995). The *g*-loadings of these four sub-tests range only from 0·81 to 0·87, and females do better on one of the two sub-tests with a *g*-loading of 0·87, and males on the other. There is thus no scope for the sex difference on this general factor to be appreciably greater than the difference in overall score, and no possibility that it could be as great as 4 points. Lubinski & Humphreys (1990) reported a slight female superiority in overall IQ score; from Hedges & Nowell's summary, it appears that females obtained higher scores on one of the three sub-tests that were averaged to give overall IQ, and males on the other two. But the largest difference was that favouring females. One would have to postulate an unbelievable difference in the *g*-loadings of these three tests to turn the overall female advantage into male superiority on *g*. Finally, on the latest standardisation of the DAT, females obtain higher scores than males on six of the eight sub-tests (Feingold, 1988). This could only translate into a male advantage of 4 points on any general factor if the loadings of the remaining two tests (space relations and mechanical reasoning) on this general factor were orders of magnitude greater than those of the other six tests. But validity studies of the DAT have invariably found that essentially all its predictive validity comes from two sub-tests—verbal reasoning and numerical ability, on both of which females obtain slightly higher scores (Anastasi, 1988). If the predictive validity of IQ tests derives from *g*, as Brand (this volume) believes, the conclusion must be that females outscore males on the DAT's general factor.

The implication of this analysis is simple: contrary to the oft-repeated, but rarely documented, claim of many IQ testers that the general factor extracted from one test battery is essentially the same as the general factor extracted from another, research on sex difference suggests that different batteries yield significantly different general factors. For present purposes, little will be gained by further pursuit of the precise nature of general intelligence defined in this way. A more profitable approach suggested by the sex-differences literature would be to follow Gustafsson (1984) in identifying *g* with Cattell's Gf, or fluid intelligence. If this identification is made, and if tests such as Raven's matrices or the DAT verbal and abstract reasoning tests provide the best measure of this factor, then the answer to the original question is clear: there is no sex difference in general intelligence worth speaking of. Moreover, this conclusion did not depend on a careful selection of items that favoured one sex being balanced out by items favouring the other—for there is no evidence that Raven's matrices were constructed in this way or that there is very much variation in the pattern of sex differences across items. It is worth adding, finally, that there would then be equally little reason to believe in that other hoary explanation of greater male eminence—greater variability in general intelligence. Neither Raven's matrices nor the DAT reasoning tests reveal reliable evidence that the variance of male scores is greater than that of females (Court, 1983; Feingold, 1992).

Hedges & Nowell (1995) show that there is a consistent tendency for males to be very slightly more variable than females on most IQ-type tests, but that only on 'measures of science achievement and the vocational aptitude scales' was the difference substantial. The difference in variability in non-verbal reasoning was small, and not sufficient to produce an excess of males over females among the top 5% of the population.

Perhaps the most important and consistent case where male performance really does seem to be substantially more variable than that of females is in mathematical ability (Hyde *et al.*, 1990; Lubinski & Humphreys, 1990; Hedges & Nowell, 1995). This difference clearly does have significant implications.

The interpretation of sex differences in spatial ability

The largest, most reliable and most persistent difference between the sexes is that observed in tests of spatial ability. As already noted, tests of three-dimensional mental rotation yield a difference of some 13–14 IQ points—a difference which, unlike many others, has shown no signs of decreasing in the past 20 years (Masters & Sanders, 1993; Voyer, Voyer & Bryden, 1995). It is not surprising, therefore, that it has attracted more attention than most others, and more attempts at explanation. Unfortunately for the understanding of the concept of spatial ability, however, these attempts have concentrated on the more distant causes, rather than on providing any theoretical account of just what it is that males and females do differently when trying to solve items in spatial IQ tests. The list of distal causes that have been proposed is a long one. It includes hypotheses about different selection pressures operating on hunting males and gathering females in earlier stages of human evolution; the suggestion that high spatial ability is determined by a single recessive gene located on the X chromosome; the postulation of differences between the sexes in lateralisation of cerebral function—differences which have themselves been attributed to factors ranging from the effects of prenatal hormones, through differences in rate of maturation, to the fact that since young girls start talking sooner than young boys, more of the brain becomes devoted to linguistic function leaving less room for spatial functions. And, of course, there have been environmental hypotheses which attribute the difference to cultural stereotypes, self-images, parental attitudes, teachers' attitudes, or the opportunity to play with different types of toy (see Halpern, 1992, for a review).

It is not the present purpose to argue that the efforts that have been made to test these hypotheses have been fruitless—although many have been singularly unimpressive and, one suspects, motivated more by the desire to prove a preconceived conclusion than by any disinterested search for scientific truth. But at least some of these efforts have provided moderately convincing evidence for or against some of these suggestions. Thus the sex-linked recessive gene hypothesis has taken something of a hammering from familial studies of resemblance in spatial abilities (Scarr & Carter-Saltzman, 1982). Conversely, the observation that at least some sex differences in test scores have decreased substantially over the past 50 years (Feingold, 1988) provides strong support for an environmental explanation of those differences. However, as noted above, tests of three-dimensional mental rotation have shown no decrease in the magnitude of sex differences over the past 20 years or so, and while the determined environmentalist will obviously be able to dismiss this, it does suggest that the difference is not simply a consequence of cultural attitudes.

One of the longer-running contenders in the field of quasi-biological accounts is Levy's 'cognitive crowding' hypothesis (Levy, 1976), which suggests that male brains are more strongly lateralised than female brains—in which verbal functions have spread to the right hemisphere, thus crowding out spatial functions. Although this basic assumption has frequently been disputed, there is certainly evidence consistent with it—for example from a recent MRI study (Shaywitz *et al.*, 1995). Moreover, although there are other gaps in the chain of reasoning underlying the hypothesis, there is one observation which suggests that the sex difference in spatial ability has something to do with differences in lateralisation: in at least some studies, it has interacted with handedness. Harshman, Hampson & Berenbaum (1983) reported that although right-handed men obtained higher scores than left-handed men on the DAT space relations test, in women the difference was reversed, with the consequence that male superiority was evident only in the right-handed majority, not in the left-handed minority. They acknowledged that this interaction has not been universally found, but reported that a post-hoc re-analysis of other studies suggested that it was a reliable effect in subjects of above average IQ. As some commentators have noted (e.g. Bishop, 1990), this sort of re-analysis runs serious risks of a Type I error. But Harshman's interaction has been confirmed in two subsequently published replications (Lewis & Harris, 1990; Gordon & Kravetz, 1991) and, for what it is worth, in a study of Cambridge undergraduates (Hilliard & Mackintosh, unpublished).

This finding is no doubt both intriguing and potentially important. It is surely extremely puzzling, and does not obviously follow from Levy's original hypothesis. Since left-handedness is associated with a somewhat more bilateral representation of language, that hypothesis may predict that left-handed males should be lower in spatial ability than right-handed males. But it is not obvious why this difference should be seen only in those of above average ability. And there seems no explanation at all for the opposite pattern of results in females. More to the present point, however, results such as these, indeed this whole line of research into the distal causes of sex differences, have shed no light whatsoever on the question that a cognitive or experimental psychologist would want to ask about sex differences in spatial ability. For the existence of such differences ought to provide a valuable source of information about the nature of the cognitive processes comprising spatial ability. For a start, the wide variation in the size of the sex difference observed in different spatial tests, coupled with the observation that this variation is not simply a function of the extent to which a particular test correlates with a general factor common to all such tests (Stumpf & Eliot, 1995), implies that there must be more than one difference in underlying process, and thus that 'spatial ability' is not a unitary concept. So much may be readily accepted—and has been proposed on other grounds (Lohman, 1988). But the study of sex differences ought to be able to take one further than this, by identifying some of the ways in which the sexes differ, and thus elucidating the mechanism of spatial problem solving. One seemingly well documented suggestion has been that there is a reliable difference between the sexes in speed of mental rotation on those tasks that require such rotation for their solution—especially on more difficult versions (Kail, Carter & Pellegrino, 1979; Tapley & Bryden, 1977). But Lohman (1986) has persuasively argued that this difference is largely a by-product of a difference in accuracy when the task becomes more difficult, either because the rotation required is greater, or because the stimulus is more complex. The implication that he draws is that one locus of the sex difference is in the ability to construct and maintain a visual image. That the distinction between spatial and non-spatial reasoning is largely one of the importance of visual imagery was, of course, one of the earliest suggestions in this field (El Koussy, 1935), and although that hypothesis has not always fared well (e.g. Neisser, 1970), this seems to be because different indices of visual imagery may be measuring quite different things, not all of which are involved in spatial problem solving. The balance of evidence remains favourable (Poltrock & Agnoli, 1986).

The present general argument, however, is that differences between different groups could be used, so to say, as an independent variable to study the nature of spatial ability, rather than as a dependent variable whose putative causes must be uncovered. For example, it is not known whether the difference in average performance between men and women on tests of mental rotation is the same, in the sense of implicating the same differences in psychological processes, as that between men of above average and those of below average ability. Nor is it known whether the difference between the two sexes is the same as that between white and Japanese Americans (Nagoshi & Johnson, 1987). Such differences provide a valuable, possibly unique, and certainly little exploited method of studying the cognitive processes underlying such abilities. In that sense, group differences are a question of some scientific interest.

References

ANASTASI, A. (1988) *Psychological Testing*, 6th edn. Macmillan, New York.

ANKNEY, C. D. (1992) Sex differences in relative brain size: the mismeasure of women, too? *Intelligence*, **16**, 329–336.

BISHOP, D. V. M. (1990) *Handedness and Developmental Disorder.* Lawrence Erlbaum, Hove, Sussex.

BURT, C. L. & MOORE, R. C. (1912) The mental differences between the sexes. *J. exp. Pedagogy*, **1**, 273–284, 355–388.

CONRAD, R. (1979) *The Deaf Schoolchild.* Harper & Row, London.

COURT, J. H. (1983) Sex differences in performance on Raven's Progressive Matrices: a review. *Alberta J. educ. Res.* **29**, 54–74.

EL KOUSSY, A. A. H. (1935) The visual perception of space. *Br. J. Psychol.* **20** (Monograph Supplement).

EVANS, B. & WAITE, B. (1981) *IQ and Mental Testing.* Macmillan, London.

FEINGOLD, A. (1988) Cognitive gender differences are disappearing. *Am. Psychol.* **43**, 95–103.

FEINGOLD, A. (1992) Sex differences in variability in intellectual abilities: a new look at an old controversy. *Rev. educ. Res.* **62**, 61–84.

GARCIA, J. (1981) The logic and limits of mental aptitude testing. *Am. Psychol.* **36**, 1172–1180.

GORDON, H. W. & KRAVETZ, S. (1991) The influence of gender, handedness, and performance level on specialized cognitive functioning. *Brain and Cognition*, **15**, 37–61.

GUSTAFSSON, J. E. (1984) A unifying model of the structure of intellectual abilities. *Intelligence*, **8**, 179–203.

HALPERN, D. E. (1992) *Sex Differences in Cognitive Abilities.* Lawrence Erlbaum, Hillsdale, NJ.

HARSHMAN, R. A., HAMPSON, E. & BERENBAUM, S. A. (1983) Individual differences in cognitive abilities and brain organization, Part I: Sex and handedness differences in ability. *Can. J. Psychol.* **37**, 144–192.

HEDGES, L. V. & NOWELL, A. (1995) Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, **269**, 41–45.

HERRNSTEIN, R. J. & MURRAY, C. (1994) *The Bell Curve.* Free Press, New York.

HYDE, J. S., FENNEMA, E. & LAMON, S. J. (1990) Gender differences in mathematics performance: a meta-analysis. *Psychol. Bull.* **107**, 139–145.

HYDE, J. S. & LINN, M. C. (1988) Gender differences in verbal ability: a meta-analysis. *Psychol. Bull.* **104**, 53–69.

JENSEN, A. R. & REYNOLDS, C. R. (1983) Sex differences on the WISC-R. *Personal. individ. Diffs*, **4**, 223–226.

KAIL, R., CARTER, P. & PELLEGRINO, J. (1979) The locus of sex differences in spatial ability. *Perception and Psychophysics*, **26**, 182–186.

LEVY, J. (1976) Cerebral lateralization and spatial ability. *Behav. Genet.* **6**, 171–188.

LEWIS, R. S. & HARRIS, L. J. (1990) Handedness, sex, and spatial ability. In: *Left-Handedness: Behavioral Implications and Anomalies*, pp. 319–341. Edited by S. Coren. Elsevier, Amsterdam.

LINN, M. C. & PETERSON, A. C. (1985) Emergence and characterisation of sex differences in spatial ability: a meta-analysis. *Child Dev.* **56**, 1479–1498.

LOHMAN, D. F. (1986) The effect of speed–accuracy tradeoff on sex differences in mental rotation. *Perception and Psychophysics*, **39**, 427–436.

LOHMAN, D. F. (1988) Spatial abilities as traits, processes, and knowledge. In: *Advances in the Psychology of Human Intelligence*, Vol. 4, pp. 182–248. Edited by R. J. Sternberg. Lawrence Erlbaum, Hillsdale, NJ.

LUBINSKI, D. & HUMPHREYS, L. G. (1990) A broadly based analysis of mathematical giftedness. *Intelligence*, **14**, 327–355.

LYNN, R. (1994) Sex differences in intelligence and brain size: a paradox resolved. *Personal. individ. Diffs*, **17**, 257–271.

MACKINTOSH, N. J. & MASCIE-TAYLOR, C. G. N. (1986) The IQ question. In: *Personality, Cognition and Values*, pp. 77–131. Edited by C. Bagley & G. K. Verma. Macmillan, London.

MASTERS, M. S. & SANDERS, B. (1993) Is the gender difference in mental rotation disappearing? *Behav. Genet.* **23**, 337–341.

NAGOSHI, C. T. & JOHNSON, R. C. (1987) Cognitive ability profiles of Caucasian vs. Japanese subjects in the Hawaii Family Study of Cognition. *Personal. individ. Diffs*, **8**, 581–583.

NEISSER, U. (1970) Visual imagery as process and as experience. In: *Cognition and Affect*, pp. 159–178. Edited by J. S. Antrobus. Little Brown, Boston.

POLTROCK, S. E. & AGNOLI, F. (1986) Are spatial visualization ability and visual imagery ability equivalent? In: *Advances in the Psychology of Human Intelligence*, Vol. 3, pp. 255–296. Edited by R. J. Sternberg. Lawrence Erlbaum, Hillsdale, NJ.

RAVEN, J. (1981) *Raven Manual Research Supplement 1.* Oxford Psychologists Press, Oxford.

REYNOLDS, C. R., CHASTAIN, R. L., KAUFMAN, A. S. & MCLEAN, J. E. (1987) Demographic characteristics and IQ among adults: analysis of the WAIS-R standardisation sample as a function of the stratification variables. *J. School Psychol.* **25**, 323–342.

ROSE, S., KAMIN, L. J. & LEWONTIN, R. C. (1984) *Not in Our Genes.* Penguin, Harmondsworth, Middlesex.

RUSHTON, J. P. & ANKNEY, C. D. (1995) Brain size and cognitive ability: correlations with age, sex, social class and race. *Psychonomic Bull. Rev.* **3**, 21–36.

SCARR, S. & CARTER-SALTZMAN, L. (1982) Genetics and intelligence. In: *Handbook of Human Intelligence*, pp. 792–896. Edited by R. J. Sternberg. Cambridge University Press, Cambridge.

SHAYWITZ, B. A., SHAYWITZ, S. E., PUGH, K. R., CONSTABLE, R. T., SKUDLARSKI, P., FULBRIGHT, R. K., BRONEN, R. A., FLETCHER, J. M., SHANKWEILER, D. P., KATZ, L. & GOSE, J. C. (1995) Sex differences in the functional organisation of the brain for language. *Nature*, **373**, 607–609.

STUMPF, H. & ELIOT, J. (1995) Gender related differences in spatial ability and the *K* factor of general spatial ability in a population of academically talented students. *Personal. individ. Diffs*, **19**, 33–45.

TAPLEY, S. M. & BRYDEN, M. P. (1977) An investigation of sex differences in spatial ability: mental rotation of three-dimensional objects. *Can. J. Psychol.* **31**, 122–130.

TERMAN, L. M. (1916) *The Measurement of Intelligence.* Houghton Mifflin, Boston.

TERMAN, L. M. & MERRILL, M. A. (1937) *Measuring Intelligence.* Houghton Mifflin, Boston.

VOYER, D., VOYER, S. & BRYDEN, M. P. (1995) Magnitude of sex differences in spatial ability: a meta-analysis and consideration of critical variables. *Psychol. Bull.* **117**, 250–270.

WECHSLER, D. (1939) *The Measurement of Adult Intelligence.* Williams & Wilkins, Baltimore.