Bias and sampling error in sex difference research

By: Douglas Wahlsten

Wahlsten, D. Bias and sampling error in sex differences research. Behavioral and Brain Sciences, 1988, 11, 214. (Commentary on paper by C.P. Benbow.)

Article:

Benbow argues that, because she can find no environmental differences to explain a small sex difference in Scholastic Aptitude Test-Mathematics (SAT-M) scores in a highly select sample of volunteers from the top 3% of seventh-grade children, we should therefore look to biology for an explanation. The reasoning here is reminiscent of arguments about racial differences in intelligence used by Jensen (1969), who declared unabashedly (and prematurely) that "Compensatory education has been tried and it apparently has failed" (p. 2) and then used this to support his assertion that differences must be largely genetic. [See also Jensen: "The Nature of the Black/White Difference on Various Psychometric Tests" *BBS* 8(2) 1985.] However, there is no necessary relationship between the ease of modifying a characteristic by changing the environment and the extent of influence by chromosomal or genetic variation. Any evidence for a biological source of a sex difference in SAT-M scores must come from well-controlled studies of the biology of the children.

Among the four kinds of physiological correlates discussed in Benbow's target article, not one provides consistent evidence of a causal link with a sex difference in mathematical ability. For allergies, there is no sex difference in the Study of Mathematically Precocious Youth (SMPY) sample. For myopia, SMPY girls are even more extreme than boys. Concerning prenatal hormones, the only indicator in the SMPY sample is birth order, which is an exceedingly indirect "measure" of physiology and is strongly related to social processes. Likewise, hand preference is not a physiological measure at all, and it is sensitive to cultural influences. Benbow's discussion of physiology is indeed "speculative." Her conclusion shows that she is not sufficiently critical of some very weak research.

For example, Benbow gives credence to two studies relating the size of the corpus callosum (CC) to sex and laterality differences. One study (deLacoste-Utamsing & Holloway 1982) claiming that females have a larger splenium (posterior portion) of the CC was based on only 9 male and 5 female brains at autopsy, and the other (Witleson 1985) claiming that the whole CC is smaller in right-handed people examined 42 deer(cancer patients. However, Nasrallah et al. (1986) failed to find any relation between CC size and sex or handedness using nuclear magnetic resonance imaging with 41 normal adults. Five other studies failed to find a sex difference in the hum CC (Bell & Variend 1985; Byne et al. 1986; Demeter et al, Oppenheim et al. 1987; Weber & Weis 1986).. Seven studies with a total sample size of over 113 male and 108 female brains have failed to replicate the results for females in the widely cited deLacoste-Utamsing and Holloway (1982) study, although results for male brains have been more consistent. The 5 female brains in the 1982 study had a mean cross-sectional area of the splenium of the CC equal to 218.3 square mm, whereas combined data from three other studies involving female brains measured in a similar way yielded a 99% confidence interval for the mean area of the female splenium from 161.3 to 178.52 square mm. The sex difference in the splenium reported in *Science* in 1982 was truly a case of sampling error.

Benbow strongly asserts that sex differences in SAT-M scores do not result from different experiences. Because human boys and girls are treated differently from the cradle to the grave, the only way Benbow can prove her point is to know in considerable detail those features of the environment that are in fact most relevant to the

nurturance of mathematical ability and how they function jointly. Then she can compare the experiences of hors and girls fairly and rigorously.

However, the SMPY project does not do this. Instead, it compares boys and girls on several rather indirect measures of experience one at a time. This approach is not likely to detect relevant features of the environment. To see this problem, let us suppose there are more than 20 different aspects of early' experience, each of which can augment or impair later mathematical ability by a small amount. Now measure the correlation between SAT-M score and an indirect indicator of each element considered separately. By fragmenting the totality of relevant experience and attenuating each effect by indirect measurement, correlations will generally be low and undetectable.

What if we were to use the same approach in a genetic study of SAT-M scores? Suppose there are more than 20 autosomal genes, each of which exerts a small influence on the development of mathematical ability and each of which is linked 1((located near) a detectable marker gene, such as one specifying some protein in the blood. Now determine what alleles cad person has at each marker locus, and then look for association between each genotype and test score. Only rarely will any significant and replicable association be found. This is why attempts to identify genes relevant to complex behaviors o abilities using genetic linkage analysis (Ashton 1986; Sturt & McGuffin 1985) or DNA restriction fragment length polymorphisms (Ellis 1986) show so little promise. A series separate tests of indirect measures of many small environmental influences on mathematical ability applied to an extremely narrow range of boys and girls would not be expected to yield much in the way of significant sex differences, so it comes as no surprise that most results of the SMPY study concerning environment are negative. In view of this, surely our current ignorance about the nature of experience cannot justify support for a biological view by default.