

Familial Studies of Intelligence: A Review

Thomas J. Bouchard, Jr. and Matthew McGue

Familial Studies of Intelligence: A Review

Abstract. A summary of 111 studies identified in a survey of the world literature on familial resemblances in measured intelligence reveals a profile of average correlations consistent with a polygenic mode of inheritance. There is, however, a marked degree of heterogeneity of the correlations within familial groupings, which is not moderated by sex of familial pairing or by type of intelligence test used.

In 1963, Erlenmeyer-Kimling and Jarvik (1) published a summary of the world literature on IQ correlations between relatives. Their finding that the pattern of correlations averaged over independent studies was consistent with the pattern predicted by a polygenic theory of inheritance has been widely cited as strong evidence for some genetic determination of IQ (2). Although the accumulation of a great many new data along with the discrediting of Burt's important study on monozygotic twins reared apart (3) has outdated that review, the authors' summary or slightly modified versions of it (4) continue to be widely reproduced (5). Recently Plomin and DeFries (6) have reported a comparison of those summary data with the results of several large, recent familial studies of IQ. They conclude that, in general, the recent studies show less resemblance between relatives than do the data reported by Erlenmeyer-Kimling and Jarvik. Their summary is not comprehensive, however, and it does not identify the factors that distinguish the two bodies of data. Roubertoux and Carlier (7) have also published a recent review, but it contains only 37 percent of the studies to be cited here.

The purpose of this report is to provide a comprehensive contemporary summary of the world literature on the IQ correlations between relatives. We have updated the 1963 summary, adding recent data and deleting several studies included in the earlier review that do not meet our methodological criteria for inclusion. Although the pattern of averages reported in this and earlier reviews is remarkably consistent with polygenic theory, the individual data points are quite heterogeneous. Therefore, we have also assessed the extent to which the reported correlations are heterogeneous and have attempted to identify some factors that contribute to this heterogeneity.

In our survey of the literature we found 140 studies that reported on familial resemblances in broad cognitive ability. These were reduced to 111 by the application of explicit selection criteria (8-12). The 111 studies, which include 59 reported in the 17 years subsequent to the Erlenmeyer-Kimling and Jarvik summary, yielded 526 familial correlations, based upon 113,942 pairings. Figure 1

displays the correlations between relatives, biological and adoptive, in the 111 studies. The median correlation in each distribution is indicated by a vertical bar. The small arrow indicates the correlation that would be predicted by a genetic model with no dominance, no assortative mating, and no environmental effects. Researchers do not subscribe to such a simple model, but it provides a noncontroversial pattern against which to compare the results of various familial groupings. Different investigators will undoubtedly fit different models to the data.

In general, the pattern of average correlations in Fig. 1 is consistent with the pattern of correlations predicted on the basis of polygenic inheritance. That is, the higher the proportion of genes two family members have in common the higher the average correlation between their IQ's.

The data set contains considerable heterogeneity, as indicated by the χ^2 statistics. In an attempt to identify the factors contributing to the heterogeneity, we subdivided the familial groupings into opposite-sex and same-sex pairings (Fig. 2) and male and female pairings (Fig. 3). Among dizygotic twins the IQ's of same-sex twins are more similar than those of opposite-sex twins. This may reflect a social-environmental effect (parents may treat same-sex twins more similarly than opposite-sex twins). The difference between nontwin same-sex and opposite-sex siblings and between same-sex and opposite-sex parent-offspring pairings is trivial. The male-female comparison does not yield consistent trends. For example, the average correlations are larger in male twins than in female twins, but the reverse is true for other siblings. The absence of any demonstrable sex effect is consistent with a polygenic theory of inheritance that does not posit the existence of sex linkage. Environmental theories that emphasize the importance of sex-role effects on general cognitive development are not supported by these results (13).

Another possible source of heterogeneity is the intelligence test used. There are many tests that purport to measure intelligence, and they may not be highly interrelated. We found great diversity in test selection. For example, the 34 corre-

lations for monozygotic twins reared together were based upon results from 22 different tests, the 41 dizygotic twin correlations upon results from 25. We do not have sufficient data to determine whether the magnitude of the familial correlation is moderated by the specific test used. We did investigate whether individually administered tests and group-administered tests produced different correlations. For the monozygotic twins reared together the 24 correlations calculated on group tests produced a weighted average of .86, and the 10 calculated on individual tests a weighted average of .84. For the dizygotic twins reared together, the weighted average of 32 correlations based on group tests is .60 and of 9 correlations based on individual tests is .61. In neither case did the distinction between group and individual test produce an appreciable effect.

The 34 correlations reported on 4672 monozygotic twin pairs reared together produce a weighted average correlation of .86. This value is very close to those reported in earlier reviews and is approximately the same for male and female pairs. Although the test of homogeneity yields a significant χ^2 value ($P < .02$), the degree of heterogeneity is not ex-

treme and can be seen to result from a few rather low correlations (14). The two most extreme values are the .58 reported by Blewett (15) and the .62 reported by Nichols (16). In both cases the sample sizes are small (26 and 36 pairs, respectively). The observation that 79 percent of the reported correlations lie above .80 convincingly demonstrates the remarkable similarity of monozygotic twins.

After deleting the Burt data we are left with test results on but 65 pairs of monozygotic twins reared apart, as reported in three separate investigations. The weighted average of .72 is much less than that found for the monozygotic twins reared together, the difference suggesting the importance of between-family environmental differences. At the same time, the magnitude of this correlation would be difficult to explain on the basis of any strictly environmental hypothesis.

Three studies give midparent-midoffspring correlations, the weighted average of these being .72. In this case the genetic expectation would depend upon the number of offspring that define the midoffspring value, and is thus indeterminate. The correlation between midparent and individual offspring does have a determinate simple genetic expectation,

which is .707. The observed weighted average of .50 is substantially less than that, a discrepancy we discuss later.

The weighted average of the 41 reported correlations in dizygotic twin pairs is .60, considerably larger than for nontwin siblings. Same-sex dizygotic pairs show somewhat greater similarity than opposite-sex dizygotic pairs (.62 versus .57), males being slightly more similar than females (.65 versus .61). As with the monozygotic twins, the test of homogeneity yields a significant value ($P < .01$), although 75 percent of the correlations fall within the narrow range between .50 and .70. The two extreme values were reported in old studies on rather small samples, the lowest being the .21 reported by Wingfield (17) in 1928 on 26 pairs, the highest the .87 reported by Merriman (18) in 1924 on 51 pairs. The greater similarity of dizygotic twins than of other siblings is most often interpreted as a reflection of greater environmental similarity. It is also likely that bias in the recruitment of dizygotic twins for study is in the direction of increasing psychological similarity (19).

The weighted average for siblings reared together is .47, which although close to the simple expectation of .50 is

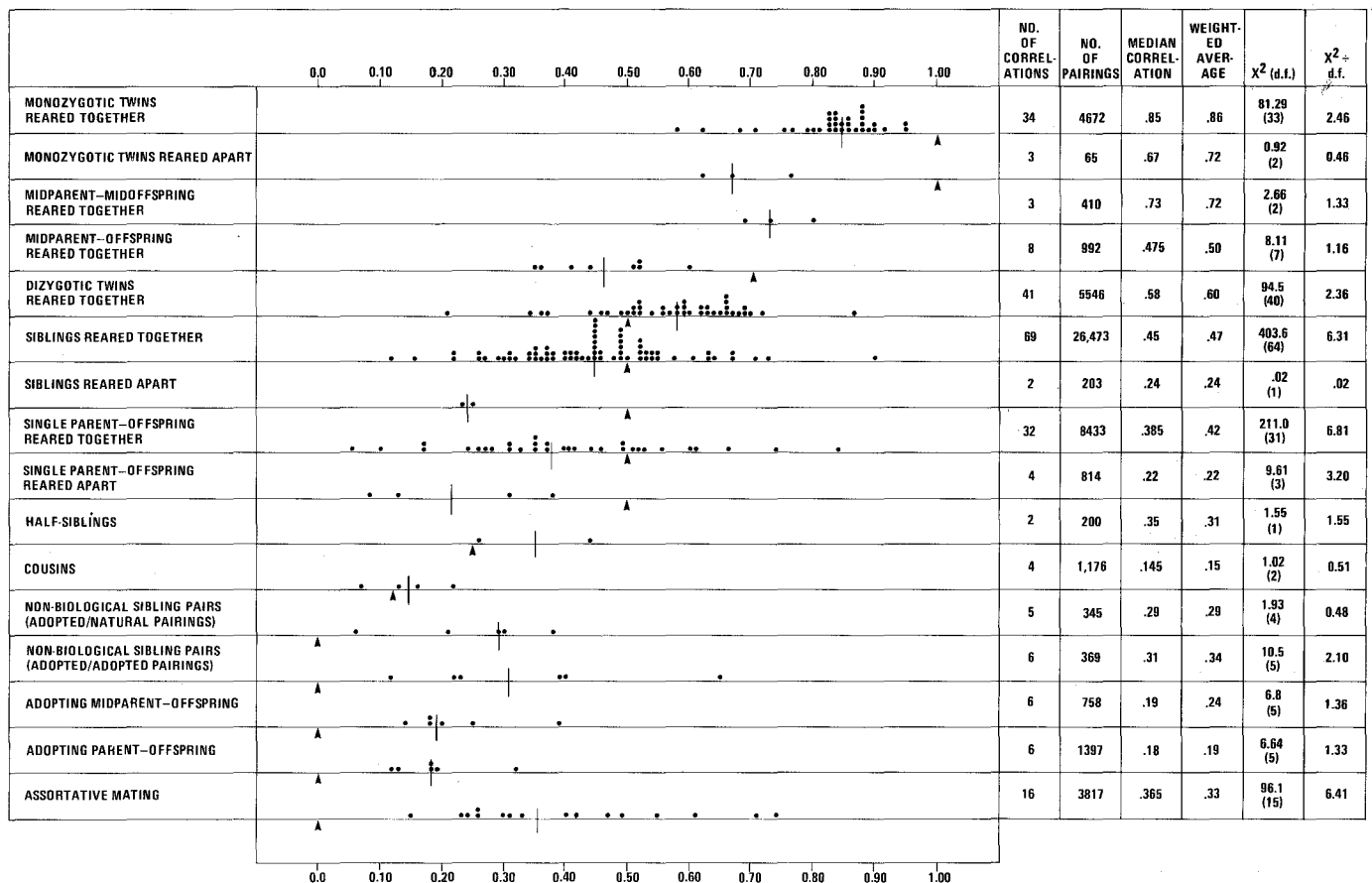


Fig. 1. Familial correlations for IQ. The vertical bar in each distribution indicates the median correlation; the arrow, the correlation predicted by a simple polygenic model.

based upon 69 values with a range of correlations from .13 to .90. Opposite-sex and same-sex siblings yield almost identical weighted averages (.49 versus .48), as do female and male siblings (.50 versus .47). The sibling correlations are based on over 25,000 pairs; one large

representative study by Record *et al.* showed a correlation of .55 on over 5000 pairs (20).

Whereas there is a wealth of information on siblings reared together, there is a dearth of information on siblings reared apart. Only 203 such pairs have been

studied, in two investigations yielding a weighted average of .24, much less than the expected value for such pairs and the average value for siblings reared together.

The weighted average correlation between individual parent and individual

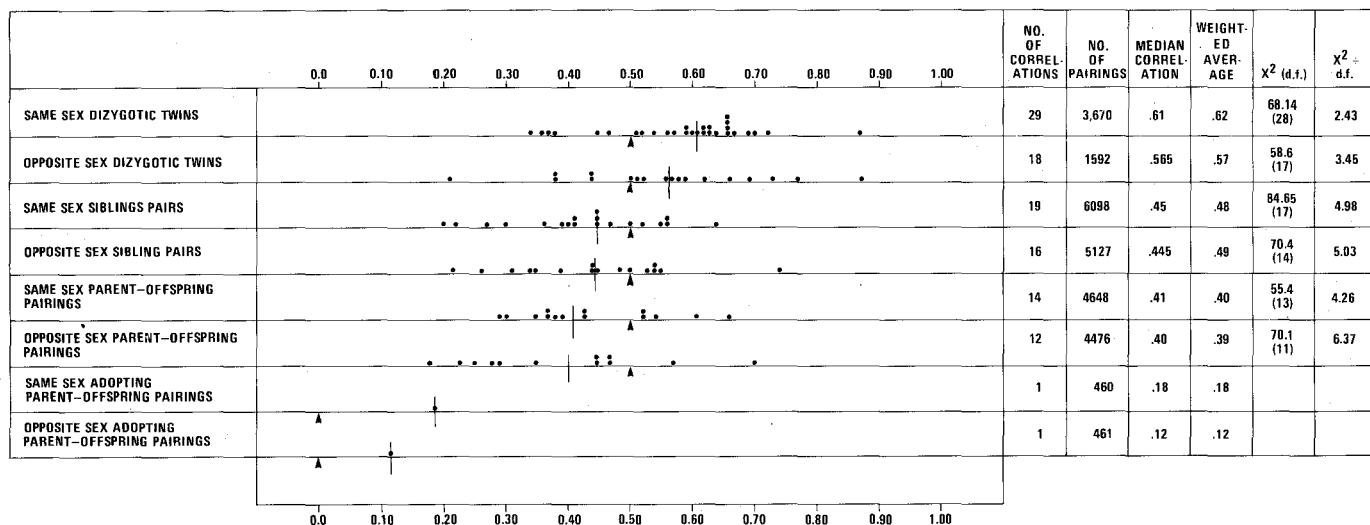


Fig. 2. Familial correlations for IQ organized by opposite-sex and same-sex pairings.

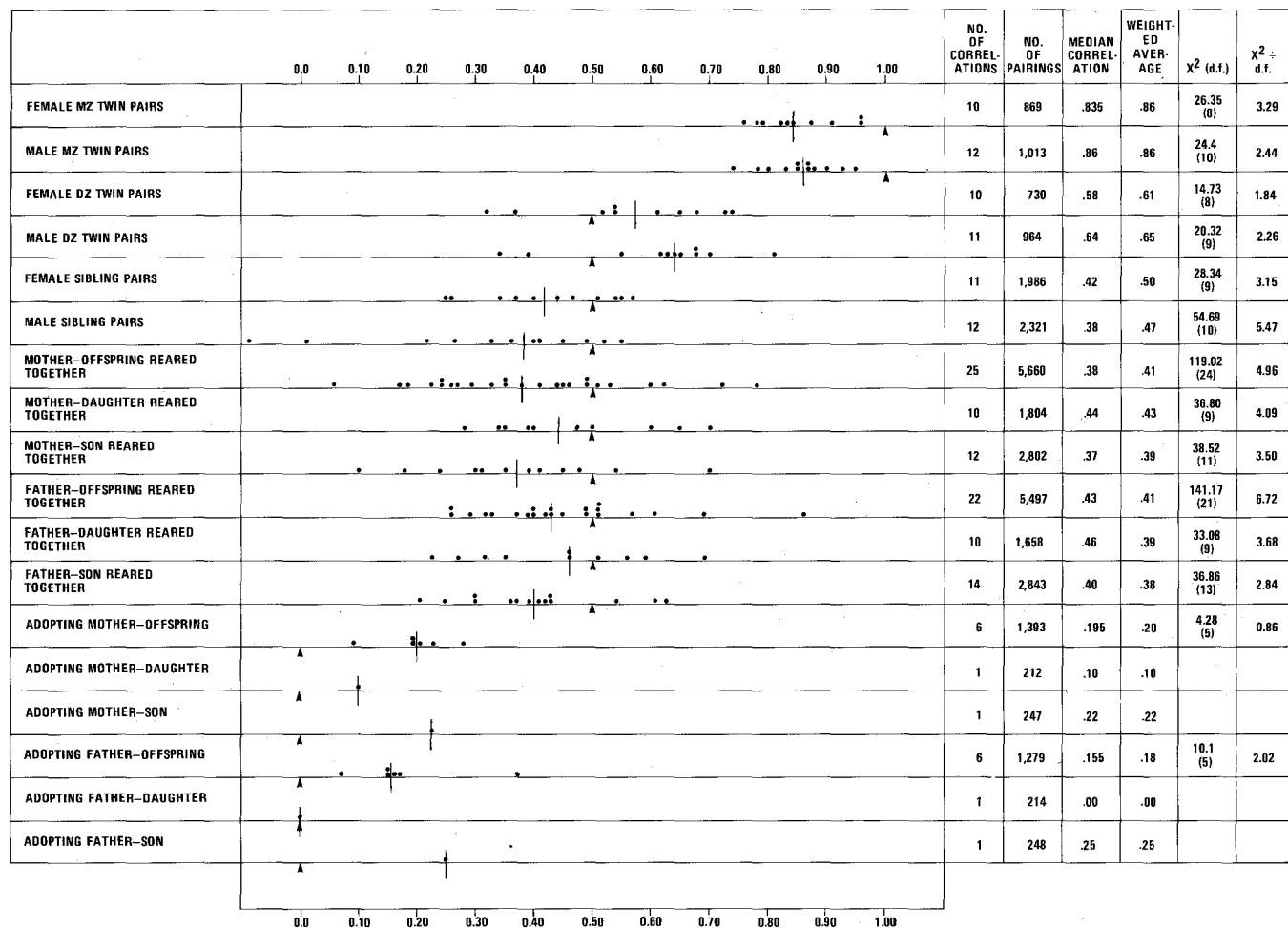


Fig. 3. Familial correlations for IQ organized by male and female pairings.

offspring is .42 based upon 32 correlations. There is a marked degree of heterogeneity in the distribution, as evidenced not only by a significant χ^2 value ($P < .01$), but also by the broad range of the correlations. Their extreme heterogeneity cannot be attributed to a sex effect, inasmuch as opposite-sex and same-sex pairings yield equivalent averages, or to a maternal effect, the average correlation of mother and offspring being the same as that of father and offspring. Although the large discrepancies between expected and observed correlations for parent and offspring reared in the parental home ("reared together") may be easily interpreted as a result of a generational (social-environmental) effect, one should not hastily discount the possibility of biological factors. Characteristics that are affected very little by the social environment, such as height and total fingerprint ridge count, show similar generational differences. For example, in one large study of height Wingerd *et al.* (21) found a midparent-offspring correlation of .51. Two large studies of total fingerprint ridge count (22) yield a Z-weighted mean midparent-offspring correlation of .63. The single parent-offspring correlations for height and total fingerprint ridge count are .42 and .42 (21-23).

As with parent and offspring reared together, correlations for parent and separated offspring are quite heterogeneous. The weighted average is .22, much less than the simple expectation of .50. As suggested by McAskie and Clarke (24), one possible explanation could be that parents and offspring are not given the same test. In fact, roughly 50 percent of our intergenerational correlations were based upon data from cases in which parents were given different tests from those given their offspring. The lower-than-expected morphological correlations found in these studies, however, suggest that scaling may not be the only problem.

Two familial pairings that are rarely studied are half-siblings and cousins. Two half-sibling correlations, both reported by Nichols (16), produce a weighted average of .31. The four reported correlations for cousins are quite homogeneous; their average, .15, closely approximates the simple genetic expectation.

A number of recent adoption studies have added considerable knowledge. Enough studies are available to permit comparison of two sets of nonbiological sibling pairs—adopted/natural and adopted/adopted. Other things being

equal, the adopted/natural correlation should be higher than the adopted/adopted correlation, because the former would contain a component for the covariance of genotype and environment (25). The present review finds the reverse (Fig. 1).

The weighted average correlation of adoptive midparent and offspring is .24, and that of adoptive parent and offspring is .19. Genetic theory requires the biological midparent-offspring correlation to exceed the biological single parent-offspring correlation, and it does, although not by much (.50 versus .42). Some environmental theories predict the same effect (24); the failure to find any difference in the adoptive case must be considered surprising from an environmental point of view.

Unlike the case in natural families, adopted offspring are somewhat more similar to the same-sex adoptive parent than to the opposite-sex adoptive parent (Fig. 2). This conclusion is based, however, on a single study (26). Overall, adoptive mothers are no more like their adopted children than adoptive fathers are.

The last row in Fig. 1 gives assortative mating coefficients. There is marked similarity between mates, but the weighted mean of .33 is much smaller than the .50 sometimes reported (27). The marked heterogeneity of the distribution indicates the sample-specific nature of these indices.

As in the earlier review, the pattern of averaged correlations is remarkably consistent with polygenic theory. This is not to discount the importance of environmental factors; monozygotic twins reared apart are far from perfectly correlated, dizygotic twins are more similar than other biological siblings, and adoptive parents' IQ's demonstrate a consistent relation with the IQ's of their adopted offspring. Although the data clearly suggest the operation of environmental effects, we found no evidence for two factors sometimes thought to be important—sex-role effects and maternal effects. That the data support the inference of partial genetic determination for IQ is indisputable; that they are informative about the precise strength of this effect is dubious. Certainly the large amount of unexplained variability within degrees of relationship, while not precluding attempts to model the data, suggests that such models should be interpreted cautiously.

THOMAS J. BOUCHARD, JR.
MATTHEW MCGUE

Psychology Department, University
of Minnesota, Minneapolis 55455

References and Notes

1. L. Erlenmeyer-Kimling and L. F. Jarvik, *Science* 142, 1477 (1963).
 2. J. D. Matarazzo, *Wechsler's Measurement and Appraisal of Adult Intelligence* (Williams & Wilkins, Baltimore, 1972).
 3. L. Hershaw, *Cyril Burt, Psychologist* (Cornell Univ. Press, Ithaca, N.Y., 1979); D. D. Dorfman, *Science* 201, 1177 (1978).
 4. L. F. Jarvik and L. Erlenmeyer-Kimling, in *Psychopathology of Mental Development*, J. Zubin and G. A. Jarvis, Eds. (Grune & Stratton, New York, 1967).
 5. P. E. Vernon, *Intelligence, Heredity and Environment* (Freeman, San Francisco, 1979).
 6. R. Plomin and J. C. DeFries, *Intelligence* 4, 15 (1980).
 7. P. Roubertoux and M. Carlier, *Ann. Biol. Clin.* 36, 101 (1978).
 8. The reports that met the following criteria were selected: (i) The research was reported in the published literature or as a doctoral dissertation. (ii) Familial resemblance was indicated by a correlation coefficient, or the information provided permitted the determination of the correlation. In excluded studies results were typically reported in terms of *F* ratios or within-pair variances. (iii) The cognitive measure employed was a widely accepted, standardized measure of intelligence, or such a measure could be derived from the test battery (for example, the first principal component as a representative of the *g* factor). In most studies excluded on the basis of this criterion either a single measure of special ability or a subjective rating of intellectual level was used. (iv) In twin studies, the procedure for zygosity determination was both objective and valid. Use of a validated questionnaire was considered an acceptable procedure. Many of the early twin studies depended on subjective procedures, for example teacher ratings. In such cases only the opposite-sex twins (necessarily dizygotic) were included in our data. (v) The sample reported on did not overlap with another sample reported elsewhere. For example, the sample reported on by Jones (9) was not taken because it is included in the larger sample reported on later by Conrad and Jones (10). (vi) Because of the controversy surrounding the samples reported on by Burt and some of his colleagues (3), all data based on them were excluded.
- For each selected study the reported correlations were classified into the relevant categories of familial relationship. If for any single sample several correlations were recorded within the same category, a single index of association was determined as follows: (i) If correlations were reported for several measures of intelligence one of which was an individually administered test, the correlation for the individually administered assessment was selected. If all measures were group intelligence tests, their average correlation was used. Averaging was on the Fisher *Z* transformations of the correlations. (ii) If several correlations were reported in a longitudinal sequence, then the correlation recorded at the highest age level was selected. (iii) If statistical corrections were applied (correction for attenuation, correction for range restriction, correction for age, or the like), only the age-corrected correlation was used. If no age-corrected correlation was reported, then the uncorrected correlation was used. (iv) If several dependent samples reported on in the same study were relevant for the category, then the weighted average of the correlations was used (weighted average of *Z* transformations, with sample size used as the weight). For example, few studies reported a single correlation between parent and offspring; instead separate father-offspring and mother-offspring correlations were given. In such cases the parent-offspring correlation was determined as the average of the two correlations with the number of pairings equal to the sum of the pairings for the two correlations.
- Application of the above rules resulted in the inclusion of only independent correlations within any single classification category. For each category the total number of pairings reported is actually the minimum number of pairings, as in several studies results were reported in terms of the number of families used. The weighted average correlation was determined by taking the weighted averages of the *Z* transformations with the inverse of the variance used as the weight and is thus the maximum likelihood estimator (11). The weights employed were $N - 3$ for interclass correlations and $N - 3/2$ for intraclass correlations. The χ^2 statistic can be used to test the hypothesis of homogeneity; that is, all

sample correlations were drawn from the same population (12). The χ^2 statistic divided by its degrees of freedom has an expectation equal to 1.0 under the homogeneity hypothesis and can be used to compare the relative heterogeneity of different categories.

Detailed information about the data presented here and a list of studies included and excluded are available from the authors.

9. H. E. Jones, *Yearb. Natl. Soc. Study Educ.* 27, 61 (1928).
10. H. S. Conrad and H. E. Jones, *ibid.* 39, 97 (1940).
11. C. R. Rao, *Linear Statistical Inference and Its Applications* (Wiley, New York, 1973).
12. M. A. G. Niana, *J. Educ. Stat.* 5, 83 (1980).
13. J. A. Sherman, for example [J. A. Sherman, *Sex-Related Cognitive Differences* (Thomas, Springfield, Ill., 1978), p. 40], argues: "Of the sex-role characteristics mentioned the most relevant to cognitive development would appear to be the emphasis on achievement and independence in problem-solving for males rather than females." Such a pattern of socialization would be expected to reduce cross-sex correlations relative to same-sex correlations.
14. With samples as large as those reported here the χ^2 test for heterogeneity is quite sensitive and could detect even trivial amounts of variability.
15. D. B. Blewett, *J. Ment. Sci.* 100, 922 (1954).
16. P. L. Nichols, thesis, University of Minnesota (1970); S. H. Broman, P. L. Nichols, W. A. Kennedy, *Preschool IQ, Prenatal and Early Developmental Correlates* (Erlbaum, Hillsdale, N.J., 1975).
17. A. H. Wingfield, *Twins and Orphans: The Inheritance of Intelligence* (Dent, London, 1928).
18. C. Merriman, *Psychol. Monogr.* 33, (No. 152) (1924).
19. D. T. Lykken, A. Tellegen, R. DeRubeis, *Soc. Biol.* 25, 1 (1978).
20. R. G. Record, T. McKeown, J. N. Edwards, *Ann. Hum. Genet.* 33, 61 (1969).
21. J. Wingerd, I. L. Solomon, E. J. Schoen, *Pediatrics* 52, 555 (1973).
22. S. B. Holt, *The Genetics of Dermal Ridges* (Thomas, Springfield, Ill., 1968); E. Matsuda, *Jpn. J. Hum. Genet.* 17, 293 (1973).
23. W. E. Nance, *Acta Genet. Med.* 25, 100 (1976).
24. M. McAskie and A. M. Clarke, *Br. J. Psychol.* 67, 243 (1976).
25. R. Plomin, J. C. DeFries, J. C. Loehlin, *Psychol. Bull.* 84, 309 (1977).
26. J. M. Horn, J. C. Loehlin, L. Willerman, *Behav. Genet.* 9, 177 (1979).
27. A. R. Jensen has reviewed this literature and found a weighted mean of .42 [A. R. Jensen, in *Human Variation*, R. T. Osborne, C. E. Noble, N. Weyl, Eds. (Academic Press, New York, 1978)]. His review is less selective than ours.
28. This research was supported in part by the National Institute of Mental Health traineeship grant 5 T32 MH14647 and a research grant from the University of Minnesota Computer Center.

22 August 1980; revised 13 January 1981