From A. R. Jensen (1970). Behaviour Genetics, *Vol 1, No. 2*, 133–146, *by kind permission of the author and Plenum Publishing Corporation*

IQ's OF IDENTICAL TWINS REARED APART

Arthur R. Jensen

Institute of Human Learning

University of California, Berkeley

ABSTRACT—A new analysis of the original data from the four largest studies (Newman, Freeman and Holzinger, 1937; Shields, 1962; Juel-Nielsen, 1965; Burt, 1955) of the intelligence of monozygotic twins reared apart, totaling 122 twin pairs, leads to conclusions not found in the original studies or in previous reviews of them. Statistical analysis of the twin differences reveals no significant differences among the twin samples in the four studies; all of them can thus be viewed statistically as samples from the same population. They can therefore be pooled for more detailed and powerful statistical treatment.

The 244 individual twins' IQ's are normally distributed, with the mean $= 96.82$, $SD = 14.16$. The mean absolute difference between twins is 6.60 ($SD = 5.20$), the largest difference being 24 IQ points. The frequency of large twin differences is no more than would be expected from the normal probability curve. The overall intraclass correlation between twins is .824, which may be interpreted as an upper-bound estimate of the heritability (h^2) of IQ in the English, Danish, and North American Caucasian populations sampled in these studies. The absolute differences between twins (attributable to nongenetic effects and measurement error) closely approximate the chi distribution; this fact indicates that environmental effects are normally distributed. That is, if $P = G + E$ (where P is phenotypic value, G is genotypic value, and E is environmental effect), it can be concluded that for this population P, G, and E, are each normally distributed. There is no evidence of asymmetry or of threshold conditions for the effects of environment on IQ. The lack of a significant correlation ($r = -0.15$) between twin-pair means and twin-pair differences indicates that magnitude of differential environmental effects is not systematically related to intelligence level of twin pairs.

Comparison of monozygotic (MZ) twins reared apart is conceptually the simplest method of estimating the broad heritability of a characteristic. Theoretically, the characteristic's total phenotypic variance (V_P) in the population is analyzable into a genetic component (V_G), a nongenetic (or "environmental") component (V_E), a component attributable to the covariance of genotypes and environments (V_{GE}), a component due to the interaction (i.e., the non-additive effects) of genetic and environmental factors (V_I), and a variance component due to measurement error (V_e). Thus:

$$V_P = V_G + V_E + V_{GE} + V_I + V_e.$$

Heritability in the broad sense is defined as $h^2 = V_G/V_P$, or, if corrected for attenuation (errors of measurement), as $h_c^2 = V_G/(V_P - V_E)$.

The correlation between pairs of individuals can be expressed as the proportion of the variance components that the members of each pair have in common:

$$r = \frac{\text{Sum of Variance Components in Common}}{\text{Total Variance}}$$

In an idealized experiment to estimate h^2, therefore, we would assign each member of a pair of genetically identical individuals to different environments entirely at random at the moment of conception, and then determine the correlation between the pairs at some later stage of their development. Since the environmental conditions are randomized there would be no correlation between pairs due to environmental effects and there would be no correlation between genotypes and environments, at least at the outset. (Different genotypes can influence the environment differently, thereby producing some genotype X environment covariance. This component is usually regarded as part of the genetic variance in heritability studies of socially conditioned characteristics.) V_G, therefore, is the only component our idealized pair would share in common, and so the correlation between them would be equal to $V_G/V_P = h^2$.

The closest approximation to this idealized experiment in reality is the study of MZ twins separated soon after birth, or in infancy and early childhood, and reared separately. Unfortunately, in such studies there is always some uncertainty about the degree to which the nongenetic variance components are common to the separated twins. There is little, if any, real doubt in the major studies about the genetic component. Errors in the determination of zygosity in these studies are highly improbable. Any such errors, of course, would subtract from V_G and thus would result in a lower value of h^2. The nongenetic components are much more questionable. There is never truly random assignment of separated twins to their foster homes. Some separated twins are reared, for example, in different branches of the same family. And twins put out for adoption rarely go into the poorest homes. Furthermore, separated twins have the same mother prenatally, and to whatever extent there are favorable or unfavorable maternal conditions that might affect the twins' intrauterine development, these conditions are presumably more alike for twins than for singletons born to different mothers. On the other hand, twin correlation due to common nongenetic factors is counteracted to some unknown extent by effects occurring immediately after fertilization which create inequalities in the development of the twins. Darlington (1954) points to nuclear, nucleocytoplasmic, and cytoplasmic differences occurring in the first stages of cell division that would cause MZ twins to be less alike than their genotype at the moment of fertilization. Some of these conditions of embryological asymmetry do not affect singletons or dizygotic (DZ) twins. Partly for this reason DZ twins are more alike in birth weight than MZ twins. Although the biologic discordances referred to by Darlington affect only a minority of MZ twins, he concludes that their total effect is sufficient to lead to a gross underestimate in all twin studies of the force of genetic determination.

The correlation between MZ twins reared apart, therefore, cannot be taken at its face value as the most valid estimate of h^2. It must be checked against estimates of h^2 obtained by other means which involve more complex formulas (and often additional assumptions) for estimating heritability from a variety of kinship correlations, including unrelated children reared together and the comparison of correlations for MZ and DZ twins. Estimates of h^2 from MZ twins reared apart are, so to speak, cross-validated when similar values of h^2 are found by other methods, assuming that similar biases do not operate in the same direction or that they are statistically controlled. There is, in fact, quite substantial agreement among the various methods and types of data for estimating heritability. Using practically all the appropriate data to be found in the literature, heritability estimates for intelligence are distributed about an average value of close to 0.8 (Jensen, 1969). MZ twins reared apart show a correlation of similar magnitude for intelligence.

The questions posed by the present study are: do the major researches on MZ twins reared apart show consistency with one another in estimates of the heritability of intelligence? Are the main parameters of these samples sufficiently alike to permit the data from the several studies to be analyzed as a total composite that would allow new and stronger inferences than would be possible for any one of the studies viewed by itself?

METHOD

The published literature contains only four major studies of the intelligence of MZ twins reared apart (Newman *et al.*, 1937; Shields, 1962; Juel-Nielsen, 1965; Burt, 1966). There are a few single sets of separated MZ twins scattered in the literature, but they are either psychiatric cases or do not present adequate intelligence test data for the purpose of the present analysis. The four major studies, based on twins from the Caucasian populations of England, Denmark, and the United States, comprise a total of 122 sets of MZ twins separated early in life and reared apart. Details concerning the twins' sex, age of separation, environmental circumstances, case histories, and so on, are to be found in the original publications. The present analysis is based on the individual intelligence test scores of the 244 subjects.

The data

Burt (1966). The 53 pairs in Burt's sample were obtained largely from schools in London. All had been separated at birth or during their first six months of life. Their IQ's were obtained from an individual test, the English adaptation of the Stanford–Binet, with mean = 100, *SD* = 15.

Shields (1962). The 44 pairs in Shield's sample were adults obtained from all parts of the British Isles. (One twin was found as far away as South America.) All of Shields' twins were separated before 6 months of age and 21 of the pairs were separated at birth. Complete intelligence test scores were obtained on only 38 of the 44 sets of twins. Two tests were used: Raven's Mill Hill Vocabulary Scale (a synonyms multiple-choice test), and the Dominoes (D48) test (a timed twenty-minute nonverbal test of intelligence). The Dominoes test has a high *g* loading (.86) and correlates .74 with Raven's Progressive Matrices. Since Shields presented the results of these tests in the form of raw scores, it was necessary to convert them to the standard IQ scale. A raw score of 19 on the Vocabulary scale and of 28 on the Dominoes Test correspond to IQ 100 in the general population. The raw score means were transformed in accord with these population IQ values and the standard deviation was transformed to accord with the population value of *SD* = 15. The IQ's thus obtained on each test were then averaged to yield a single IQ measure for each subject.

Newman, et al. (1937). These 19 twin pairs were obtained in the United States and were tested as adults. In 18 cases the age of separation was less than 25 months, and in 9 it was less than 6 months. About the one pair that was separated at 6 years (and tested at age 41) Newman *et al.,* states: ". . . the twins were separated at six years, somewhat late for our purposes; but we had information that the environments of the twins had been so markedly different since separation that we decided to add the case to our collection" (p. 142). (These twins differed by 9 IQ points.)

Stanford–Binet IQ's were obtained on all subjects.

Juel-Nielsen (1965). These 12 pairs were obtained in Denmark. The age of separation ranges from 1 day to 5¾ years; 9 were separated before 12 months. IQ's were obtained by an individual test, a Danish adaptation of the Wechsler-Bellevue Intelligence Scale (Form I), which in the general population has a mean = 100 and *SD* = 15.

TABLE 1

IQ's for MZ Twins Reared Apart

Burt (1966), $N = 53$ Pairs									
A	B	A	B	A	B	A	B	A	B
68	63	94	86	93	99	115	101	104	114
71	76	87	93	94	94	102	104	125	114
77	73	97	87	96	95	106	103	108	115
72	75	89	102	96	93	105	109	116	116
78	71	90	80	96	109	107	106	116	118
75	79	91	82	97	92	106	108	121	118
86	81	91	88	95	97	108	107	128	125
82	82	91	92	112	97	101	107	117	129
82	93	96	92	97	113	108	95	132	131
86	83	87	93	105	99	98	111	......	......
83	85	99	93	88	100	116	112	......	......
Shields (1962), $N = 38$ Pairs*									
95	87	109	102	102	108	76	79	84	68
96	100	98	110	113	111	91	84	121	121
95	79	101	87	89	93	103	116	107	111
71	75	99	108	88	110	98	94	74	69
86	84	99	97	96	99	94	76	79	84
105	105	69	71	85	84	95	101	107	106
93	76	86	85	89	84	96	97	......	......
83	89	107	105	90	107	63	73	......	......
Newman *et al.* (1937) $N = 19$ Pairs									
85	97	89	93	102	96	94	95	105	115
78	66	94	102	122	127	84	85	96	77
99	101	105	106	116	92	90	91	79	88
106	89	77	92	109	116	88	90	......	......
Juel-Nielsen (1965) $N = 12$ Pairs									
120	128	100	94	99	105	114	124	......	......
104	99	111	116	100	94	114	113	......	......
99	108	105	97	104	103	112	100	......	......

*IQ's transformed from raw scores on Mill Hill Vocabulary tests and the Domino D48 Test. (See text for explanation.)

IQ's of Identical Twins Reared Apart

The IQ's of all the twins in the four studies are given in Table 1.

RESULTS

The main statistical parameters of the separate studies and of the combined data are shown in Table 2. The few instances of slight discrepancies between these statistics and the corresponding figures of the original authors are all within the range of rounding error. All the present analyses were calculated by computer, with figures carried to five decimals and not rounded till the final product.

TABLE 2

Statistics on IQ's of MZ Twins

Study	*N* (Pairs)	Mean IQ	*SD*	$\|d\|$	SD_d	r_i	r_d
Burt	53	97.7	14.8	5.96	4.44	.88	.88
Shields	38	93.0	13.4	6.72	5.80	.78	.84
Newman *et al.*	19	95.7	13.0	8.21	6.65	.67	.76
Juel-Nielsen	12	106.8	9.0	6.46	3.22	.68	.86
Combined	122	96.8	14.2	6.60	5.20	.82	.85

Distribution of IQ's

The mean IQ of the MZ twins is slightly below the population mean. This is a general finding for twins reared together or apart and is probably related to the intrauterine disadvantages of twinning, including lowered birth weight. The small Juel-Nielsen sample is atypical in having a mean IQ above 100. The stand-

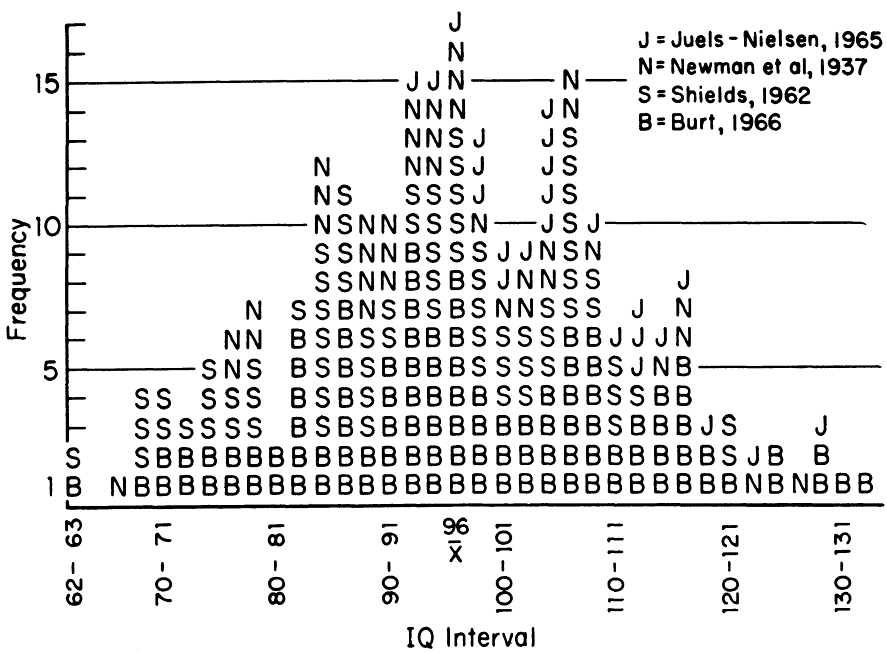

FIGURE 1. IQ distribution of 244 MZ twins reared apart, from four studies. The distribution does not deviate significantly from normality.

ard deviation of the twin IQ's is only slightly less than the 15 points in the general population. Figure 1 shows the form of the IQ distribution. It extends over a range of 71 IQ points, or 4.7 sigmas, which would include approximately 98 percent of the general population. A chi square test of the goodness of fit shows that the IQ distribution of Figure 1 does not depart significantly from normality. The chi square based on eight subdivisions of the distribution is only 3.08, $p = 0.80$. (Chi square with 7 degrees of freedom must exceed 14.07 for significance at the .05 level.) It can be concluded that the IQ's of the total sample of 244 twins are quite typical and representative of the distribution of intelligence in the general population.

Correlation between twins

The intraclass correlations (r_i) between twins are given in Table 2. A correlation scatter diagram for all twins is shown in Figure 2. Twins were assigned to the *A* and *B* axes in such a way as to equalize the means of the two distributions. The intraclass correlation (r_i) represented by the scatter diagram is .82. Corrected for attenuation (i.e., test unreliability), assuming the upper-bound for Stanford–Binet test reliability of .95, the twin correlation would be .86.

It is interesting to compare the scatter diagram for IQ's shown in Figure 2 with a scatter diagram for the socioeconomic status (*SES*) of the homes in which

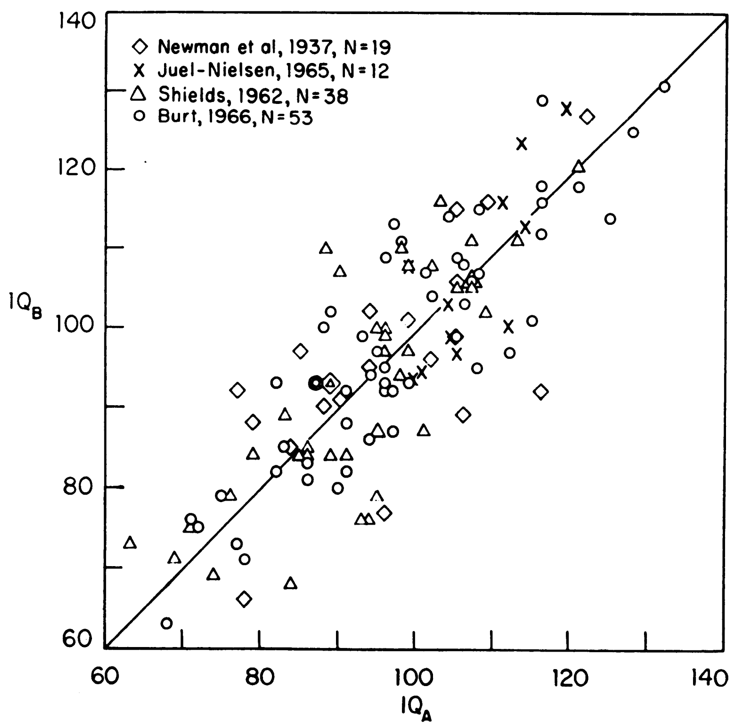

FIGURE 2. Scatter diagram showing correlation between IQ's of 122 sets of co-twins (A and B assigned at random). The obtained intraclass correlation (r_i) is 0.82. The diagonal line represents perfect correlation ($r_i = 1.00$).

the twins were reared. The one study which classified subjects in terms of *SES*, based on parents' or foster parents' occupation, is Burt's. The six categories were (1) higher professional, (2) lower professional, (3) clerical, (4) skilled, (5) semi-skilled, (6) unskilled. The seven cases reared in residential institutions are omitted from this analysis, since there is no basis for assignment to one of the six *SES* categories. The scatter diagram is shown in Figure 3. It represents a correlation of 0.03 between the SES of the homes of the separated twins in Burt's

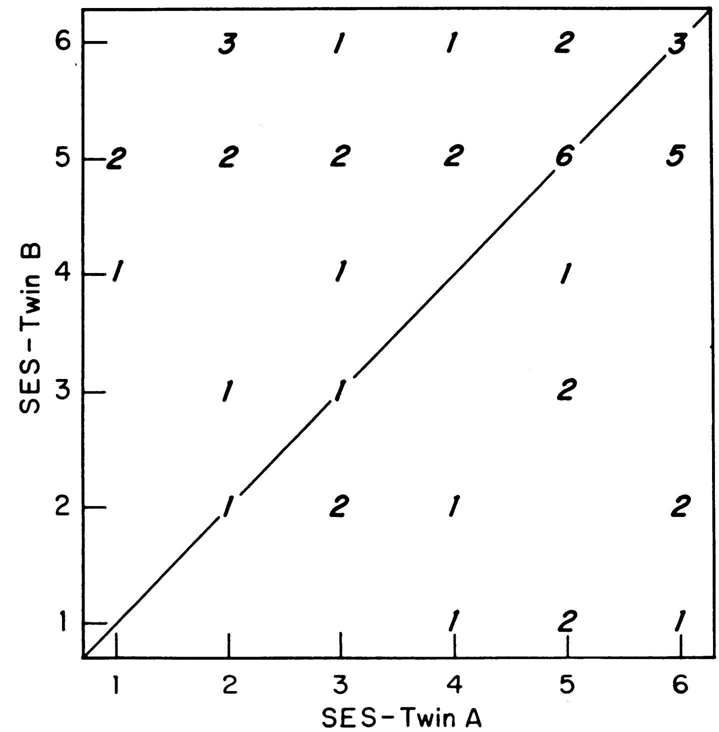

FIGURE 3. Scatter diagram of socioeconomic status (SES, based on six occupational categories of the parents, from "professional" (#1) to "unskilled" (#6)) for 46 co-twins in the Burt (1966) study. The numbers in the scatter diagram represent frequencies of twin-pairs. (Assignment to A and B is the same as in Figure 2.) The intraclass correlation (r_i) between co-twins' SES is 0.03.

sample. Obviously virtually none of the correlation between twins' IQ's is attributable to similarities in their home environments when these are classified by *SES* in terms of the parents' occupation.

The intraclass correlations for IQ in the four studies differ from one another mainly because of differences in the restriction of range of IQ's in the various samples. The magnitude of r_i is, of course, partly a function of the sample variance. The magnitude of r_i by itself, therefore, can be a somewhat deceptive indicator of the actual magnitude of twin differences (or similarities) relative to the population variance. For this reason the most crucial statistic in twin data is the absolute difference between twins.

Twin differences in IQ

The mean absolute difference ($|\bar{d}|$) between twins and the standard deviation of the differences (SD_d) are shown in Table 2. Since the absolute difference between twins also contains measurement error due to imperfect reliability of the tests, the $|\bar{d}|$ of 6.60 should be compared to the value of 4.68, which is the mean difference between forms *L* and *M* of the Stanford–Binet administered to the same persons. The *SD* of these differences is 4.13 (Terman and Merrill, 1937, p. 46). Some of this difference, of course, reflects gains due to the practice effect of the first test upon the second. But the mean difference of 6.60 can be corrected for attenuation assuming the upper bound reliability for the Stanford–Binet of .95, which results in a "true" absolute difference of 5.36

It is proposed that the absolute differences between twin's IQ's can be used to compute a correlation coefficient which has the same scale as the Pearson and intraclass correlation but indicates the degree of similarity between twins relative to the similarity between persons paired at random from the general population. This can be called a "difference correlation," signified as r_d. This is a useful statistic in studying kinship resemblance because it preserves the actual magnitude of the difference between kinship pairs. For example, even if there were a perfect Pearson r (or intraclass correlation) between relatives, r_d would be less than 1.00 if there was any mean difference between the related persons (as would be the case if one member of each pair of MZ twins were reared in a very unfavorable environment and one member in a very favorable environment). Thus r_d should be reported in twin studies (and other kinship studies) to supplement the usual correlation coefficient (Pearson or intraclass). The value of r_d is not sensitive to the sample variance. Imagine that by some fluke we obtained a sample of twins with no differences between the means of the twin pairs; even if the average difference between members of each pair were small, the intraclass correlation (or Pearson r) between twins would be zero, suggesting that the heritability is zero. Especially when twin samples are small, it makes more sense to ask what is the magnitude of the twin differences relative to differences among unrelated persons in the general population. The answer is provided by r_d. The formula for r_d is

$$r_d = 1 - \left(\frac{|\bar{d}_k|}{|\bar{d}_P|}\right)^2 ,$$

where

$|\bar{d}_k|$ = mean absolute difference between kinship members,

$|\bar{d}_P|$ = mean absolute difference between all possible paired comparisons in the general population, and

$$|\bar{d}_P| = \frac{2\sigma}{\sqrt{\pi}} = 1.13\sigma.$$

Unless one has an estimate of σ in the population from which the kinship groups are a sample or to which one wishes to generalize concerning r_d, this statistic cannot be used.

IQ's of Identical Twins Reared Apart

It can be seen in Table 2 that the values of r_d are much more consistent than r_i among the four studies. Corrected for attenuation (reliability = .95) the composite r_d of .85 becomes .88. This value should be interpreted as an estimate of h^2 only with caution, since it is uncertain just how much of the nongenetic variance is common to the separated twins. The studies do not differ significantly in r_d, because the values of $|\bar{d}|$ themselves do not show significant differences among the studies. An analysis of variance to test the significance of differences in $|\bar{d}|$ in the four studies yielded an $F = 0.87$, $df = 3$ and 118, $p < 0.46$. Thus the studies clearly do not differ significantly in the magnitude of twin differences. Bartlett's test was performed on the standard deviations of the absolute differences (SD_d) and revealed that on this parameter the differences among the studies are nonsignificant at the .01 level.

Figure 4 shows the frequency distribution of the absolute differences between twins. These are, of course, composed of environmental effects plus errors of

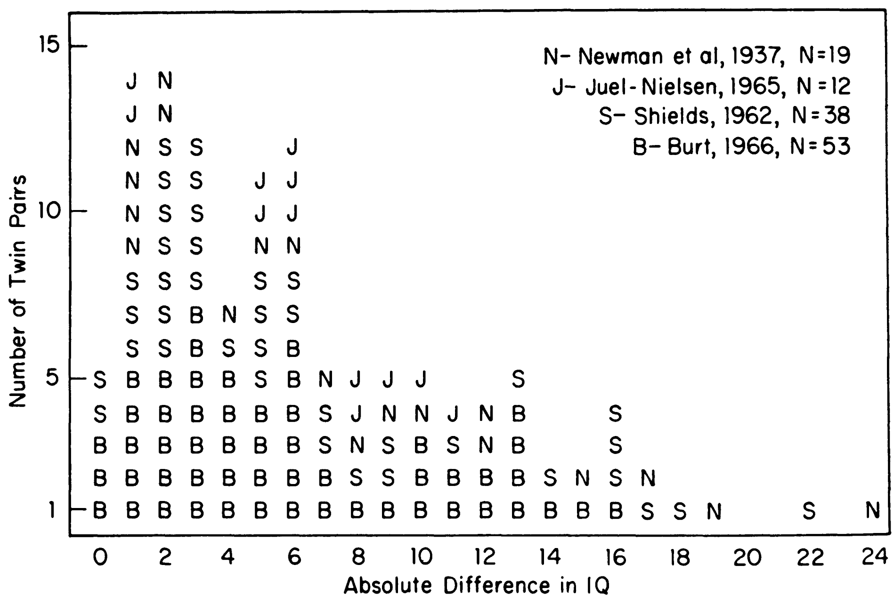

FIGURE 4. Distribution of absolute differences ($|d|$) in IQ between co-twins reared apart. This distribution closely approximates the chi distribution.

measurement. Extreme differences are rare; in only 3 cases does $|\bar{d}|$ exceed the average difference of 17 IQ points between all possible pairs of persons in the population; and in only 19 cases (16 percent) do the differences exceed the average difference of 12 IQ points between full siblings reared together, while 16 percent of the differences exceed the mean difference of about 11 IQ points generally found between DZ twins reared together. Since the differences shown in Figure 4 represent environmental effects (and random errors of measurement), these results should permit some inference about the distribution of environmental effects on IQ.

Distribution of environmental effects

The distribution of absolute differences shown in Figure 4 closely resembles a chi distribution. If one draws pairs of values at random from a normal distribution, the absolute differences between the values in each pair yield the chi distribution, which, in effect, is one half of the normal distribution. One can think of the chi distribution as consisting of the normal distribution folded over on itself, with the fold at the median. (The corresponding deviations above and below the median, of course, are added together.) Therefore, one can graphically test a distribution for goodness of fit to the chi distribution by plotting the obtained distribution on a normal probability scale after the percentiles of the distribution have been "unfolded" at the median. This "unfolding" is simply achieved by the transformation 50 + %ile/2. If these values when plotted on the normal probability scale fall approximately along a straight line, it is evidence that the distribution does not differ significantly from chi. Figure 5 shows

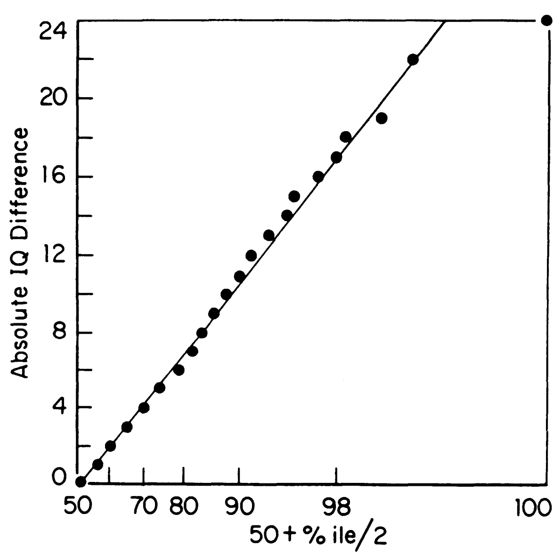

FIGURE 5. The absolute differences in IQ between co-twins plotted against a normal probability scale. The close fit to the straight line shows that environmental effects on the IQ, as represented by co-twin differences, are normally distributed.

this plot. The goodness of fit of the data to a straight line is practically perfect with the exception of the one most extreme case among the 122 twin pairs—an IQ difference of 24 points. This is the frequently cited case of Gladys (IQ 92) and Helen (IQ 116) in the study by Newman *et al.* (p. 245). They were separated at 18 months and tested at the age of 35 years. They had markedly different health histories as children; Gladys suffered a number of severe illnesses, one being nearly fatal, while Helen enjoyed unusually good health. Gladys did not go beyond the third grade in school, while Helen obtained a B.A. degree from a good college and became a high school teacher of English and history.

What Figure 5 means is that the nongenetic or environmental effects, which are wholly responsible for the twin differences, are normally distributed. (The absolute differences are due to environmental effects plus measurement error; it is assumed that errors of measurement are distributed normally.) Note that this says nothing about the distribution of environments *per se*. The conclusion refers to the *effects* of environment on IQ. There is no evidence in these data of asymmetry or of threshold conditions for the effects of environment on IQ.

Since the IQ's (i.e., phenotypes) are themselves normally distributed (Figure 1), and since the environmental effects on IQ have been shown to be normally distributed in this sample, it follows that the genotypes for IQ also are normally distributed. (The sums of two normal variates also have a normal distribution.) That is to say, if $P = G + E$ (where P is phenotypic value, G is genotypic value, and E is environmental effect), it can be concluded that for these IQ data, P, G, and E are each normally distributed.

Since P, G, and E are distributed normally, it is meaningful to estimate the standard deviations of their distributions. (We assume test reliability of .95 and normally distributed errors of measurement.) Given these conditions and a twin correlation (r_d) of .85, the estimates that would obtain in a population with $\sigma =$ 15 are shown in Table 3. Since in a normal distribution six sigmas encompass

TABLE 3

Components of Variance in IQ's Estimated from MZ Twins Reared Apart

Source	σ	σ^2	% Variance
Heredity	13.83	191.25	85
Environment	4.74	22.50	10
Test Error	3.35	11.25	5
Total (Phenotypes)	15.00	225.00	100

virtually 100 percent of the population (actually all but 2 x 10^{-7} percent), and since the standard deviation of environmental effects on IQ is 4.74, it can be said that the total range of environmental effects in a population typified by this twin sample is 6 x 4.74 = 28.4 I.Q. points.

Genotype X environment interaction

A corollary to the finding that environmental effects are normally distributed is the question of whether a favorable environment raises the IQ more or less than an unfavorable environment depresses the IQ. If favorable and unfavorable environmental effects were asymmetrical, we should expect to find that the higher and lower IQ's from each pair of twins would have different distributions about their respective means. This is in fact not the case. Probably the way to see this most clearly is to plot the IQ's of the higher and lower twins in each pair against the absolute difference between the twins. This plot is shown in Figure 6. The mean IQ's of the higher and lower twins are 100.12 and 93.52, respectively. The difference is significant beyond the .001 level. The corresponding *SD*'s are 13.68 and 13.86; the difference is nonsignificant. The straight lines through the data points are a least squares best fit. The slopes of these lines (in opposite directions) are not significantly different. The correlation (Pearson r) between IQ and

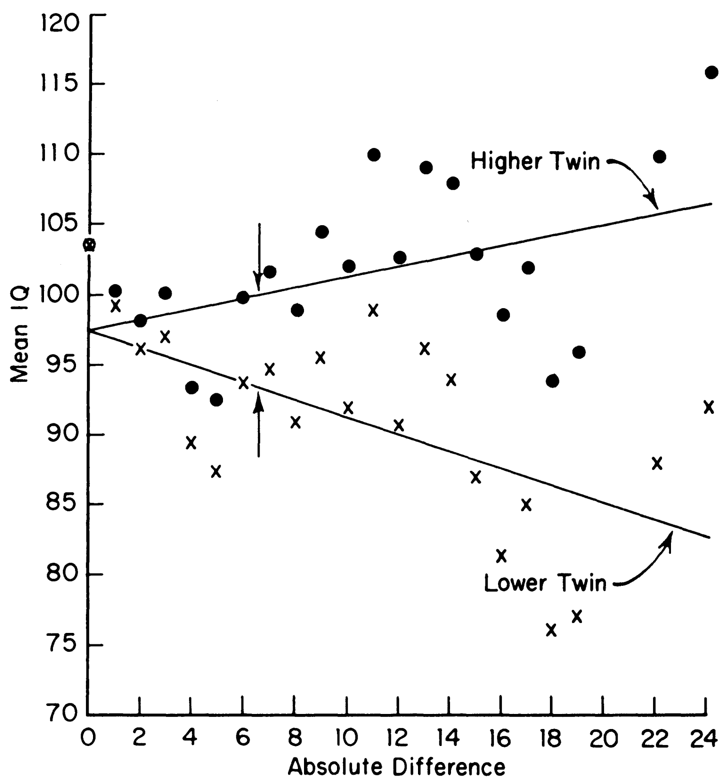

FIGURE 6. IQ of the higher twin (H) and the lower twin (L) plotted against their absolute difference in IQ. The straight lines are a least squares best fit to all the data (122 twins). The straight arrows indicate the bivariate means.

absolute difference is + 0.15 for the lower twins and – 0.22 for the higher twins. The difference (disregarding the sign of *r*) is completely nonsignificant.

We can also ask: Is there an interaction between environment and genotype for intelligence? If there is, we should expect a correlation between the mean IQ of each twin pair (reflecting their genotypic value) and the absolute difference between the twins (reflecting environmental differences).[1] This correlation (Pearson *r*), based on the 122 pairs, turns out to be – 0.15, which is not significantly different from zero. These data, then, do not show evidence of a genotype X environment interaction for IQ.

Sources of environmental differences

The present data do not permit any strong inferences about the sources of environmental variance, but other twin research indicates that a substantial and perhaps even a major proportion of the nongenetic variance is attributable to prenatal and other biological influences rather than to differences in the social-psychological environment. The cytoplasmic discordances and the like pointed

---

[1]This method of assessing the GXE interaction was originally suggested and explicated by J. L. Jinks and D. W. Fuller in "Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior." *Psychological Bulletin*, 1970, 73, 311–49.

out by Darlington have already been mentioned. Differences in the favorableness of the intrauterine environment are reflected in differences in birth weight between twins (the differences being greater for MZ than for DZ twins), and the differences in birth weight are known to be related to IQ disparities in twins. In a review of this evidence, Scarr (1969) found that MZ twins who were both over 2500 grams in birth weight differed in later IQ by 4.9 points in favor of the heavier twin; when one twin was less than 2500 grams, the IQ difference was 13.3; and when both twins were less than 2500 grams, the IQ difference was 6.4. The mean difference of 6.9 IQ points between the heavier and lighter MZ twins (52 pairs) in the studies summarized by Scarr is not far from the mean IQ difference of 6.6 between all the twins in the present study.

It is sometimes argued that the IQ resemblance between MZ twins reared apart is largely attributable to similarities in their home environments. To the extent that this is true, it should lead to the prediction that characteristics with *lower* heritability (and consequently greater susceptibility to environmental influences), should show even less difference between MZ twins reared apart, as compared with MZ twins reared together, than characteristics of *higher* heritability. In this connection it is instructive to compare the IQ with tests of scholastic achievement for MZ twins reared together and reared apart. A review of studies of the heritability of scholastic achievement has shown much lower values of h^2 (the average being about 0.40) than for IQ (Jensen, 1967). The studies by Burt and Newman *et al.* provide the necessary scholastic achievement data for the relevant comparisons. These are shown in Table 4. Note that when twins are reared together (MZT), they differ much less in scholastic achievement than when

TABLE 4

Mean Absolute Difference ($|\bar{d}|$) Between MZ Twins Reared Together (MZT) and Reared Apart (MZA) for IQ and Scholastic Achievement (Both scaled to $\sigma = 15$)

	IQ		Sch. Ach.		Number	
Study	MZT	MZA	MZT	MZA	MZT	MZA
Burt	4.79	5.96	2.40	10.29	95	53
Newman *et al.*	5.90	8.21	3.39	11.86	50	19
Combined	5.17	6.55	2.74	10.70	145	72

reared apart (MZA). No such large difference is found for IQ between MZT and MZA. If the MZA twin resemblance in IQ were due to environmental similarities, these similarities should be even more strongly reflected by scholastic achievement, and this is clearly not the case. Estimates of *within* and *between* family environmental effects may be obtained by subtracting $(MZT)^2$ from $(MZA)^2$ and obtaining the square root. For IQ the *within* environments effect is 5.17 and the *between* environments effect is $(6.55)^2 - (5.17)^2 = \sqrt{16.2} = 4.02$ IQ points. For scholastic achievement the *within* envirnments effect is 2.74 and the *between* environments effect is 10.34. The fact of much greater *within* than *between* environmental effects for IQ strongly suggests that the differences between identical twins in IQ arise largely from prenatal factors rather than from influences in the social-psychological environment. Just the opposite conclusion would pertain in the case of scholastic achievement.

CONCLUSION

Analysis of the data from the four major studies of the intelligence of MZ twins reared apart, totaling 122 twin pairs, leads to conclusions not found in the original studies or in previous reviews of them. A statistical test of the absolute difference between the separated twins' IQ's indicates that there are no significant differences among the twin samples in the four studies. All of them can be viewed as samples from the same population and can therefore be pooled for more detailed and powerful statistical treatment.

The 244 individual twins' IQ's are normally distributed, with the mean = 96.82, *SD* = 14.16. The mean absolute difference between twins is 6.60 (*SD* = 5.20), the largest difference being 24 IQ points. The frequency of large twin differences is no more than would be expected from the normal probability curve.

The overall intraclass correlation between twins is .824, which may be interpreted as an upper-bound estimate of the heritability of IQ in the English, Danish, and North American Caucasian populations sampled in these studies.

The absolute differences between members of twin pairs (attributable to nongenetic effects and measurement error) closely approximate the chi distribution; this fact indicates that environmental effects are normally distributed. If $P = G + E$ (where P is phenotypic value, G is genotypic value, and E is environmental effect), it can be concluded that for this population P, G, and E are each normally distributed. There is no evidence of asymmetry or of threshold conditions for the effects of environment on IQ. The lack of a significant correlation between twin-pair means (reflecting genotype values) and twin-pair differences (reflecting environmental effects) indicates a lack of genotype X environment interaction; that is to say, the magnitude of differential environmental effects is not systematically related to the intelligence level of twin pairs. Additional evidence from comparison of the difference between MZ twins reared together with the difference betwen MZ twins reared apart suggests that most of the small twin difference in IQ may be attributable to prenatal intrauterine factors rather than to later effects of the individual's social-psychological environment.

REFERENCES

Burt, C. (1966). The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *British Journal of Psychology,* **57,** 137–53.

Darlington, D. C. (1954). Heredity and environment. Proc. IX International Congress of Genetics. *Caryologia,* **190,** 370–81.

Jensen, A. R. (1967). Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proceedings of the National Academy of Sciences,* **58,** 149–57.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* **39,** 1–123.

Juel-Nielsen, N. (1965). Individual and environment: a psychiatric-psychological investigation of monozygous twins reared apart. *Acta psychiatrica et neurologica Scandinavica,* (Monogr. Suppl. 183).

IQ's of Identical Twins Reared Apart

Newman, H. H., Freeman, F. N., and Holzinger, K. J. (1937). *Twins: A study of heredity and environment,* University of Chicago Press, Chicago.

Scarr, Sandra. (1969). Effects of birth weight on later intelligence. *Social Biology,* **16,** 249–56.

Shields, J. (1962). *Monozygotic twins brought up apart and brought up together,* Oxford University Press, London.

Terman, L. M., & Merrill, Maud A. (1937), *Measuring intelligence,* Houghton-Mifflin, Boston.

RÉSUMÉ—Une nouvelle analyse des données originelles tirées des quatre plus importantes études (Newman, Freeman and Holzinger, 1937; Shields, 1962; Juel-Nielsen, 1965; Burt, 1966) sur l'intelligence de jumeaux monozygotes élevés séparément et portant au total sur 122 paires de jumeaux, conduit à des conclusions qui n'apparaissent pas dans les études originelles ou leurs critiques faites jusquà présent. L'analyse statistique des différences entre les jumeaux montre qu'il n'y a pas de différence significative entre les échantillons de jumeaux des quatre études; ainsi, toutes les paires peuvent être considérées, d'un point de statistique, comme des échantillons d'une même population. Elles peuvent donc être regroupées afin de donner lieu à une analyse statistique plus poussée et plus détaillée.

Les quotients intellectuels des 244 jumeaux, pris individuellement, sont distribués suivant une loi normale, de moyenne = 96,82 et d'écart-type = 14,16. La différence moyenne, en valeur absolue, entre les jumeaux est 6,60 (écart-type = 5,20), la différence la plus grande étant égale à 24 points de quotient intellectuel. La fréquence d'occurrence de grandes différences entre jumeaux n'est pas plus grande que ce que l'on pouvait attendre en se basant sur la courbe normale de probabilité. Globalement, le coefficient de corrélation entre jumeaux à l'intérieur de chaque classe est égal à 0,824, évaluation qui peut être considérée comme au-dessus de la moyenne des estimations du degré de transmission héréditaire (h^2) du quotient intellectuel parmí les populations anglaises, danoises et causiennes de l'Amérique du Nord, d'où ont été tirés les échantillons pour effectuer ces études. Les valeurs absolues des différences entre jumeaux (attribuables à des effects non génétiques et à des erreurs de mesure) approche de très près la distribution du χ^2; ceci indique que les effects dûs à l'environnement sont distribués d'une façon normale. Cest-à-dire que si $P = G + E$ (où P est la valeur phénotype, G est la valeur génotype et E l'effet dû à l'environnement), on peut en conclure que, pour cette population, P, G, et E sont chacun normalement distribués. Rien ne laisse supposer qu'en ce qui concerne les effets de l'environnement sur le quotient intellectuel, il puisse y avoir une assymétrie ou un seuil. L'absence de corrélation significative (r = —0,15) entre les moyennes et les différences se rapportant à chaque paire de jumeaux indique que l'ampleur des effets différentiels dûs à l'environnement ne peut pas être systématiquement liée au niveau d'intelligence des paires de jumeaux.

ZUSAMMENFASSUNG—Eine erneute Auswertung der in den vier umfangreichsten Studien (Newman, Freeman and Holzinger, 1937; Shields, 1962; Juel-Nielsen, 1965; Burt, 1966) enthaltenen Daten über die Intelligenz von eineiigen Zwillingen, die getrennt erzogen worden sind (insgesamt 122 Zwillingspaare), führt zu Folgerungen, die in den ursprünglichen Studien bzw. in früheren Besprechungen derselben nicht enthalten sind. Die statistische Auswertung der Unterschiede zwischen den Zwillingen zeigt, dass zwischen den Zwillings-Auswahlgruppen in den vier Studien keine bedeutsamen Unterschiede bestehen, sodass man sie statistisch alle als Auswahlgruppen der gleichen Gesamtmasse betrachten kann. Folglich können sie zum Zwecke einer detaillierteren und überzeugenderen statistischen Behandlung zusammengelegt werden.

Die Intelligenzquotienten der 244 eineiigen Zwillinge weisen eine Normalverteilung mit dem arithmetischen Mittel = 96,82 und der Standardabweichung SD = 14,16 auf. Der mittlere absolute Unterschied zwischen Zwillingen beträgt 6,60 (Standardabweichung SD = 5,20), der grösste Unterschied beträgt 24 IQ-Punkte. Die Häufigkeit der grossen Unterschiede zwischen Zwillingen ist nicht grösser als die normale Wahrscheinlichkeitskurve erwarten lässt. Die Gesamtkorrelation innerhalb der Gattung zwischen Zwillingen ist 0,824, was als Schätzung des oberen Grenzwertes der Erblichkeit (h^2) des Intelligenzquotienten für die kaukasische Gesamtmasse in Dänemark, England und Nordamerika, deren Auswahlgruppen in

diesen Studien betrachtet wurden, gedeutet werden kann. Die absoluten Unterschiede zwischen Zwillingen (nichtgenetischen Einflüssen und dem Messfehler zuschreibbar) kommen der chi-Verteilung sehr nahe; dies bedeutet, dass die Umwelteinflüsse eine Normalverteilung aufweisen. Das heisst, wenn $P = G + E$ (P steht für den phänotypischen Wert, G für den genotypischen Wert und E für die Umwelteinflüsse), so kann geschlossen werden, dass im Falle dieser Gesamtmasse P, G und E jeweils eine Normalverteilung aufweisen. Es bestehen keine Anzeichen für das Vorliegen einer Asymmetrie oder von Schwellenwertbedingungen bezüglich der Umwelteinflüsse auf den Intelligenzquotienten. Die Abwesenheit einer bedeutsamen Korrelation ($r = -0{,}15$) zwischen den Zwillingspaar-Mittelwerten und den Zwillingspaar-Unterschieden deutet darauf hin, dass das Ausmass unterschiedlicher Umwelteinflüsse nicht systematisch mit dem Intelligenzniveau von Zwillingspaaren zusammenhängt.

Manuscript received April 27, 1970