Behavior Genetics, Vol. 4, No. 1, 1974

Kinship Correlations Reported by Sir Cyril Burt

Arthur R. Jensen[1]

Received 26 Apr. 1973–Final 22 May 1973

All the kinship correlations (and their sample sizes) reported over a period of 30 years by Sir Cyril Burt are presented in tabular form. The kinships include identical twins reared together and apart, fraternal twins, siblings, parent-child, grandparent-grandchild, cousins, and others, more rare types of relationships. Burt's statistical methods and the procedures for testing and obtaining "final assessments" of mental ability are fully described, and the final assessments for Burt's 53 monozygotic twins reared apart are given in full. Misprints and inconsistencies in some of the data are noted, and recommendations are made for the presentation and preservation of kinship data secured by future researchers.

KEY WORDS: Burt; genetics; intelligence; relatives; twins.

THE BASIC DATA

The late Sir Cyril Burt throughout his long career as a psychologist collected data on the mental resemblance between persons of various kinships, and he intermittently published reports of these findings. The reports were often, but not always, cumulative; that is to say, Burt frequently combined the more recently acquired data with those of previous reports. Much of the testing was done by Burt himself, and with the help of his research assistants, in surveys carried out in the schools of greater London during Burt's tenure as psychologist to the London County Council (see Jensen, 1972*a*). The quest for the exceedingly rare and genetically most valuable data on identic twins who were separated in infancy and reared apart, however, continued successfully into the

[1]University of California, Berkeley, California.

years after Burt's retirement from his Professorship at University College, London. In 1955, Burt commented that the quest for separated monozygotic or one-egg twins had lasted for over 40 years; data on 21 pairs had been collected up to that time (Burt, 1955, p. 167, footnote 3).

Because various kinship correlations are of considerable value to researchers in behavioral genetics who wish to test the fit of particular genetic models to empirical estimates based on mental test data, it seemed worthwhile to bring together in one set of tables all of the various kinship correlations reported throughout Burt's writings, which, on this topic, extend over a period of some 60 years. Reports of his first empirical studies of mental resemblance among relatives, however, go back only about 30 years.

THE MENTAL MEASUREMENTS

Burt's kinship studies were based on four types of mental measurements: group tests of intelligence, individual tests of intelligence, "final assessments," and tests of scholastic achievement and "general attainments," i.e., scholastic knowledge and skills. Usually he also included in his genetic analyses highly heritable physical measurements (height, weight, head length and width, and eye color) as a basis for comparison with the parallel analyses performed on the mental measurements.

Burt never gave very detailed descriptions of the specific psychological tests he used in any given study, but all of the tests he used are presented in two of Burt's books on mental and scholastic tests (Burt, 1921, 1923). Briefly, the intelligence tests are described as "(i) a group test of intelligence containing both non-verbal and verbal items, (ii) an individual test of intelligence (the London Revision of the Terman Binet [i.e., Stanford-Binet] Scale) used primarily for standardization, and for doubtful cases, (iii) a set of performance tests, based on the Pintner-Paterson tests and standardized by Miss Gaw (1925)" (Burt, 1966, p. 140).

As for the scholastic measurements, Burt states that "we used the group tests constructed and standardized for London children" (1966, p. 140). These tests are to be found in Burt (1921). So much, then, for the group and individual tests of intelligence, and the scholastic achievement tests. These are quite straightforward. However, the data that Burt seemed to put most reliance on, what he called "final assessments" or "adjusted assessments," are less clear-cut. Therefore, a close examination of what Burt's "final assessments" consisted of, insofar as one can tell from his writings, is in order.

Final Assessments

First, it should be noted that while Burt was a leading pioneer in psychometric theory and one of the world's most sophisticated workers in this field right up until the time of his death in his eighty-ninth year, in actual practice his approach to psychological testing and assessment was more that of the clinician than of the psychometric researcher. His thinking on this matter is best expressed in his own words:

> we shall be concerned, not with special abilities, but only with general intelligence. Here once again it is to be remembered that what the psychologist can directly observe or measure is only the phenotype, never the genotype. Hence our first aim has been a practical one–to estimate the relative efficiency of the methods available for assessing intelligence. We might almost borrow Galileo's explanation when he repeated Archimedes' experiment: "I am not trying to discover how many ounces of gold there may be in the crown, but simply how much accuracy there is in the method." For this purpose we have endeavoured to ascertain the comparative merits of the two most obvious procedures: first, the use of standardized tests, with the marks taken just as they stand [i.e., the group and individual tests]; secondly, checking and correcting the marks so obtained in the light of the available evidence [i.e., the "adjusted assessments"]. . . . as a means of estimating genotypic differences, even the most carefully constructed tests are highly fallible instruments, and their verdicts are far less trustworthy than the judgments of the pupil's own teachers; the outstanding merit of the tests is that, unlike the teacher's judgments, they are comparable from one school to another. Accordingly, our investigations have been directed towards discovering, first, how far such tests are likely to be affected by irrelevant influences, and secondly how far existing methods of assessment can be improved upon by more comprehensive devices. Most psychological investigators have concentrated on the first of these two questions only. Almost inevitably the investigator who enters the schools from outside is obliged to confine himself to short and simple procedures–group tests of the usual type. And here lies the unique advantage of an officially appointed psychologist who is himself a member of the education authority's staff: he not only knows at first hand what are the conditions affecting the examinees in the various schools, but also possesses the authority to extend or repeat his tests and his interviews, and to require from teachers the detailed information and the further assistance that he needs.
>
> The data we have used for the purpose of our various analyses have been secured in this way; and, having satisfied ourselves that by these means we can reduce the disturbing effects of environment to relatively slight proportions, we have gone on to enquire whether the data so obtained–the frequency curves, the bivariate distributions, the correlations between relatives brought up under various conditions (twins, cousins, children and their parents, siblings living in the same home, in different homes, in residential institutions, and the like)–are consistent with the hypothesis of multifactor inheritance. That has formed our chief theoretical problem. (Burt and Howard, 1957, pp. 38-39)

But what did these "final assessments" actually consist of? How were they arrived at? To what extent are the procedures replicable by other investigators? Unfortunately, there is no sufficiently detailed account in any of Burt's articles of precisely how his final assessments were arrived at for us to feel confident that the procedures could be precisely repeated by other investigators. This is not to say that the end results of similar procedures might not be much the same. Probably the fullest description Burt ever gave of his method of arriving at the final assessments is the following:

> To assess intelligence as we have defined the term, it will be unwise to rely exclusively on formal tests of the usual type, particularly when they consist largely of verbal group tests and have been applied on a single occasion only. Provided the

> person tested is in good health at the time of the examination, has a strong motive for doing his best, and is in no way handicapped by special disability, emotional disturbance, or a lack of the elementary knowledge which the problems presuppose, then such a test may no doubt furnish a tolerably accurate assessment. But this cannot be taken on trust; and the only way to be sure that no distorting influences have affected the results is to submit the marks to some competent observer who has enjoyed a first-hand knowledge of the testees. With children this will usually be a school teacher; and whenever discrepancies appear between the teacher's verdict and that of the test, the child must be re-examined individually, preferably with tests of a non-verbal type.
>
> Such precautions were regularly adopted by the psychologist working in the London County Council schools, since one of his main duties was concerned with the ascertainment of possible cases of certifiable deficiency and of scholarship ability which had been overlooked or questioned. The interview, the use of non-verbal tests, and the information available about the child's home circumstances usually made it practicable to allow for the influence of an exceptionally favorable or unfavorable cultural environment. Whether the effects of infantile ailments or irregular growth could always be allowed for is perhaps more questionable. In any case, when test-results have been systematically checked and adjusted by these means, the reliability of the final assessments, and their correlation with the pupils' subsequent achievements, prove to be far higher than those of a single intelligence test, whether group or individual. (Burt and Howard, 1956, pp. 121-122)

In 1958, Burt stated that "assessments are obtained by submitting the test scores to the teachers for criticism or correction, and where necessary adjusting them by the methods described above" (Burt, 1958, p. 9). Namely: "The final assessments for the children were obtained by submitting the marks from the group tests to the judgment of the teachers who knew the children best: where the teacher disagreed with the verdict of the marks, the child was interviewed personally, and subjected to further tests, often on several successive occasions. The assessments for the adult members of the family were naturally far less accurate" (Burt, 1958, p. 8).

Burt added somewhat more information on the final assessments in 1966:

> The test-results . . . were submitted to the teachers for comment or criticism; and, wherever any question arose, the child was re-examined. It was not practicable for the same person to test every child. I was helped by three principal assistants, and in a few cases by research-students, all of whom had been trained by me personally. The methods and standards therefore remained much the same throughout the inquiry. If any divergence occurred, it would tend to lower rather than to raise the correlations. The reliability [type of reliability not specified] of the group test of intelligence was 0.97; of the Stanford Binet 0.95; of the performance tests 0.87. (Burt, 1966, p. 140)

It is clear that Burt intended his final assessments to "read through" the effects of environment as much as possible. He wrote:

> Environment appears to influence the test results chiefly in three ways: (*a*) the cultural amenities of the home and the educational opportunities provided by the school can undoubtedly affect a child's performance on intelligence tests of the ordinary type, since so often they demand an acquired facility with abstract and verbal modes of expression; (*b*) quite apart from what the child may learn, the constant presence of an intellectual background may stimulate (or seem to stimulate) his latent powers by inculcating a keener motivation, a stronger interest in

> intellectual things, and a habit of accurate, speedy, and diligent work; (*c*) in a few rare cases illness or malnutrition during the prenatal or early postnatal stages may, almost from the very start, permanently impair the development of the child's central nervous system. The adjusted assessments may do much towards eliminating the irrelevant effects of the first two conditions; but it is doubtful whether they can adequately allow for the last. (Burt, 1958, p. 9)

The process Burt described for arriving at "final assessments" is exactly what a good clinician should do in arriving at judgments or decisions which can personally affect individuals. But one may question whether the subjective element in this procedure is one that can be wholeheartedly recommended in scientific research on the genetics of mental abilities. Since it is not completely explicit, it cannot be completely objective, and therefore not entirely repeatable by other investigators.

If all *S*s had been given the same set of tests, with the scores combined into a "final assessment" in some completely specifiable manner, this would be an altogether different matter, and, in my opinion, much more satisfactory. I made this point to Burt in the conversations I had with him in the summer of 1970. He said that for the vast majority of *S*s the final assessments involved no subjective judgments at all, but consisted simply of an average of the IQs obtained on the group test and the individual test, and in some cases when teachers completely agreed with the group test scores the final assessments consisted of no more than the single group test IQs.

This was not true, however, of the twins in Burt's analyses. Their greater rarity and theoretical value caused all of them to be given the "full treatment"; they were given all the tests in arriving at the final assessments, and in some cases the final assessments were even revised between one published study and the next on the basis of further testing to which they had been subjected during the interim. For example, between 1958 and 1965, a number of the twins who entered into the 1955, 1956, and 1958 studies were retested by one of Burt's principal assistants and the new scores were used to revise the final assessments on these *S*s (Burt, 1966, p. 141, footnote).

THE KINSHIP CORRELATIONS

In presenting all of the kinship correlations that appear anywhere in Burt's writings, the guiding principle has been to avoid inferences and interpretation in the tabular presentation itself but to point out in the text accompanying each table certain questions to which its contents give rise. In order to avoid inference and interpretation in my preparation of these tables, I have simply presented everything I could find in a careful search through Burt's writing,with the source of each item clearly identified. I have neither eliminated any data nor combined any data, but have endeavored to tabulate for the reader simply what exists, to serve as a completely objective basis for whatever inferences, interpretations, or uses the reader may wish to make of the basic information. This approach makes it necessary to repeat a good deal of the same information which often appears in more than one of Burt's articles. But I believe the reader will find that this form of repeated tabulation gives a clearer and much more comprehensive picture of the raw sum total of Burt's empirical legacy to behavioral genetics than would be given by a more succinct form of summary. It has the added advantage of turning up possible errors, and inconsistencies can easily go unnoticed until all the data are brought together in this fashion.

The following tables present the kinship correlations to as many decimal places as reported by Burt in the reference cited, along with the number of pairs (*N*), when this was given; a question mark (?) designates those instances where the *N* is not explicitly given in the cited reference. Typographical errors which occurred in the original articles are always presented in the body of the tables just as they occurred in the original article, and the corrected values are given in the footnotes. These corrections, in all cases, were explicitly made by Burt himself; none is inferred. Where there is a question of other possible misprints or errors, they are mentioned speculatively in the text and are kept clear of the tables and footnotes.

The measure of the degree of relationship generally used by Burt is the intraclass correlation (Burt, 1943, p. 91, footnote 4; Burt and Howard, 1956, p. 124), as advocated by R. A. Fisher (1934, p. 213), and this is correct. The usual product-moment correlation (Pearson *r*) assumes that the two correlated variables, *X* and *Y*, can be separated and assigned to two distinct classes of variable (like height and weight). Thus the Pearson *r* is sometimes referred to as the *interclass* correlation. But since there is no basis for assigning each of the members of a twin pair to one class or another, it is entirely arbitrary whether a given twin is either *X* or *Y*. And since different arbitrary assignments of twins to *X* and *Y* in computing the product-moment correlation will result in different values of *r*, this type of correlation is inappropriate. The intraclass *r* gets around this; in effect, it amounts to an unbiased estimate of the mean *r* in the population that would be obtained from all possible assignments of twins to the *X* and *Y* variables.

Karl Pearson (1901) was the first to suggest a method for obtaining the average sample *r*, using the product-moment correlation technique; he used it to compute correlations between siblings and other kinships. Burt told me in 1970 that he may at times have used Pearson's method instead of the intraclass *r*, but he clearly preferred the intraclass *r*, since, unlike Pearson's *r*, it is an unbiased estimate of ρ, i.e., the correlation in the population. The results of the two methods are practically the same when *N* is large, and they usually differ, if they differ at all, only in the third decimal place; the larger the *N*, of course, the smaller the difference between the results of the two methods. Pearson's method is merely the product-moment correlation with double entry of the *X* and *Y* variables. For example, a pair of twins (or other kins) called *A* and *B* are entered into the computations twice, first with *A* assigned to *X* and *B* to *Y*, and then with B assigned to X and A to Y. This causes the X and Y arrays to have exactly the same means and standard deviations, and under this condition the product-moment correlation is nearly equal to the intraclass correlation. But not quite, unless N is very large, because the X and Y standard deviations that enter into the product-moment correlations are computed with N in the formula, while the intraclass correlation (which is based on the analysis of variance), being an unbiased estimate, uses the degrees of freedom (df), i.e., $N-1$. If one uses N rather than df in performing the analysis of variance, the intraclass r is exactly the same as the product-moment r obtained by Pearson's double-entry method. Of course, the larger the N, the less difference it makes whether one uses N or $N-1$. Very slight variations can result, therefore, depending on the method used for calculating the correlation.[2] But a single-entry Pearson r is always wrong for kinship correlations, unless just by chance it happens that X and Y have exactly the same mean and standard deviation. Newman *et al.* (1937), in their well-known study of twins, consistently used Pearson's double-entry method.

Twins

An essential question in all studies of twins is the method for determing zygosity, i.e., whether a given pair of twins is monozygotic ("identical") or dizygotic ("fraternal"). It is commonly believed that this is a difficult determination to make, but actually for the vast majority of twins it is not. A simple checklist of ten questions, which can be filled out by the twins themselves or by an interviewer, has been found to be about 95% as accurate as the most elaborate and refined method for determining zygosity by means of blood-group analysis, which can have an error rate of less than one in a million (see Nichols and Bilbro, 1966; Bulmer, 1970). Burt's method of determining zygosity falls somewhere between these extremes. Here is his most complete description of it:

> To determine "zygosity," i.e., to distinguish "identical" from so-called "fraternal" twins, no one criterion is sufficient. A difference of sex is of course decisive; such a pair cannot be monozygotic. With the younger children, particularly those born in the area in which the investigator was working, we were frequently able to secure detailed records of the mother's pregnancy and birth; and it is generally agreed that twins reported as born in a single chorion are monozygotic. When twins brought up together can be seen side by side, the impressionistic judgment of an expert observer is likely to be correct in nineteen cases out of twenty; but a few dizygotic twins are remarkably alike. Height, weight, and right- or left-handedness are unreliable; hair-color, eye-color, fingerprints and palm prints are more helpful. In doubtful cases the most valuable check is provided by an investigation of blood

[2] An excellent brief explication of the differences between the interclass correlation and the intraclass correlation is provided by Mather (1964, Chap. X). The most extensive exposition is by Haggard (1958).

> groups and serum groups. This had not been introduced when we started our inquiries; and, though in half-a-dozen of our later cases where slight doubts existed, it was adopted as an extra precaution, we were unable to carry it out as a routine procedure. We think it highly unlikely that any misclassifications have been made; but, if they have, their effect would be to reduce the differences between the correlations for monozygotic and dizygotic pairs. (Burt, 1966, pp. 141-142)

Monozygotic Twins Reared Together

The correlations for monozygotic twins reared together are given in Table I. In corresponding with Burt about one of my own articles (Jensen, 1968) in which I reproduced (in my Table 1) correlations from Burt's 1955 and 1958 articles, Burt informed me of two errors[3] in my table, one of them due to a typographical error in his 1955 article (the correlation of 0.944 which should be 0.904) which was unwittingly transferred into his 1958 article, in which he had simply reproduced his whole Table 1 from the 1955 article. However, if the 1955 final assessment r is 0.925 for $N = 83$, does it not seem rather improbable that the 1966 (and 1972) final assessment should be exactly the same, even though the sample size has increased from 83 to 95? But I know of no satisfactory way of determining the probability of obtaining exactly the same r under such conditions. (We see the same repetition of the r for Arithmetic, despite the increased N.)

When the various twin pairs (or any other kinship pairs) in a sample differ in age, part of the correlation can be attributable to the sameness (or similarity) in age of the members of each pair. How did Burt take this into account? With IQ scores there is really no problem, since the IQ is either a ratio of mental age/chronological age, in which case chronological age is, in a sense, partialed out of the correlation, or a standard score (with mean = 100, $\sigma = 15$ within each age group), in which case age is also partialed out. But what about the physical measurements and the scholastic measures? At no point is Burt entirely clear on this. Presumably, the achievement tests, like the intelligence tests, consisted of standardized scores or age-ratio scores (it is not clear which). As for the physical measurements, Burt wrote simply: "measurements for height and weight were corrected for sex and age" (1966, p. 141). How this was done is not specified,

[3]The other error in my 1968 Table 1 is in giving $N = 30$ for monozygotic twins reared apart; it should be $N = 21$. Burt's 1958 article did not give the Ns in his Table 1, which he later informed me was transplanted whole from his 1955 article (in which the Ns are listed separately in the text). I had inferred an $N = 30$ from the following statement in the text of Burt's 1958 article, referring to monozygotic twins reared separately: "We have now collected over 30 such cases" (Burt, 1958, p. 7). Presumably at the time Burt's 1958 article was written (it was originally the Walter Van Dyke Bingham Memorial Lecture given at University College, London, on May 21, 1957), either all 30 of the twin pairs referred to had not yet been tested or the new data had not yet been completely analyzed, so Burt used the r based on $N = 21$ in his 1958 Table 1, which is exactly the same as the 1955 Table 1.

Table I. Monozygotic Twins Reared Together

	1943, p. 91		1955, p. 168 Table 1		1956, p. 124		1958, p. 6 Table 1		1966, p. 146 Table 2		1972, p. 182 Table 1		1972, p. 186 Table 3	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*[a]	*N*	*r*	*N*	*r*
Intelligence														
Group test			83	0.944[b]	?	0.904	?	0.944[b]	95	0.944				
Individual test	62	0.86	83	0.921			?	0.921	95	0.918				
Final assessment			83	0.925	?	0.925	?	0.925	95	0.925			95	0.925
Unspecified											84	0.896		
Scholastic														
Reading and spelling			83	0.944			?	0.944	95	0.951				
Arithmetic			83	0.862			?	0.862	95	0.862				
General attainments			83	0.898			?	0.898	95	0.983	84	0.898	95	0.983
Physical														
Height			83	0.957					95	0.962	84	0.939	95	0.962
Weight			83	0.932					95	0.929				
Head length			?[c]	0.963					95	0.961				
Head breadth			?[c]	0.978					95	0.977				
Eye color			?[c]	1.000					95	1.000				

[a] Figures for boys and girls have been calculated separately and then averaged.

[b] Burt stated that this is a typographical error; correct *r* is 0.904.

[c] 1955, p. 167, footnote 3 states: "The figures for head-length, head-breadth, and eye color are based on much smaller numbers in every batch."

and nothing at all in this respect is said about the measurements of head length and breadth.

The most obvious method, and that used in the twin study by Newman *et al.* (1937, p. 95), is to obtain the twin (double-entry) correlation based on the raw measurements as well as the double-entry correlation of age with the raw measurements and then to compute the partial correlation, i.e., the correlation between twins with age partialed out. This method, of course, should have no effect on age-standardized scores such as the IQ, from which the variance due to age is already removed. If it does have an effect on IQ correlations, it can only mean biased sampling of *S*s, since for a properly standardized test the correlation between IQ and age should be zero. But when *S*s are drawn from any single grade in school, for example, there is generally found a negative correlation between IQ and age, since the lower-IQ children are old for their grade in school and the higher-IQ children are young for their grade. When there has been unbiased sampling, it is unnecessary to partial out age from IQ scores; but when there has been biased sampling (as in the school-grade example), it is statistically improper and only compounds errors to partial out age from IQ data.

Though Burt was not explicit concerning how he handled age in his correlations, there is no evidence that he made improper use of partial correlations. Newman *et al.* (1937), however, do compute partial *r*s for IQs without issuing the necessary caveat that such correlations, if different from the nonpartialed *r*s, should be disregarded and certainly never be used in formulas for estimating heritability. But I rather doubt that Burt ever used the partial correlation method for handling age at all, for I recall from our discussions in connection with my own research on twin and sibling resemblance that he expressed dislike of partialing out age as a general procedure, since this method assumes linear regression of the test variable (or physical measurement) on age, an assumption which is seldom justified except within a quite limited age range. Burt clearly favored the use of standardized scores based on as narrow age groups as the size of the normative sample would reliably permit. Unfortunately, Burt did not report the ages of his various samples, which makes his findings less precisely comparable with other studies, since some kinship correlations are known to vary as a function of age (e.g., Honzik, 1957).

Since in two of his articles Burt quotes the twin correlations of Newman *et al.* for comparison with his own twin correlations, is there the possibility that we can infer whether Burt's correlations are partial *r*s (i.e., age partialed out) or simple (zero-order) *r*s by noting which of the *r*s from Newman *et al.* Burt selected for comparison with his own data? Unfortunately, this line of possible inference leads to naught, because of the peculiar inconsistencies and, at points, sheer inaccuracies, in Burt's tabulation of the correlations from Newman *et al.*, In Table II are shown the twin correlations of Newman *et al.* as Burt tabulated them in his 1955 and 1966 articles, and exactly as tabulated in the monograph

Table II. Twin Correlations from Newman *et al.* (1937) as Reported by Burt (1955, 1966) and by Newman *et al.* (1937)

Test	MZ twins reared together				MZ twins reared apart				DZ twins reared together			
	Burt		Newman *et al.*[a]		Burt		Newman *et al.*[b]		Burt		Newman *et al.*[c]	
	1955[d] N=?	1966[e] N=50	Raw r N=50	Partial r N=50	1955[d] N=?	1966[e] N=19	Raw r N=19	Partial[f] r N=19	1955[d] N=?	1966[e] N=51	Raw r N=50	Partial r N=50
Intelligence												
Group test (Otis IQ)	0.922	0.922	0.922	0.922	0.727	*0.727*	0.727	0.609	0.621	0.621	0.621	0.618
Individual test (Stanford-Binet IQ)	0.910	*0.881*	0.910	0.881	0.670	*0.767*[g]	0.670	0.653	0.640	*0.631*	0.640	0.631
Scholastic												
General attainments (Educational age[h])	0.955	*0.892*	0.955	0.892	0.507	*0.583*[i]	0.507	0.499	0.883	*0.696*	0.883	0.696
Physical												
Height	0.981	*0.932*	0.981	0.932	0.969	0.969	0.969	0.968[j]	0.930	*0.645*	0.934	0.645
Weight	0.973	*0.917*	0.973	0.917	0.886	0.886	0.886	0.764[k]	0.900	*0.631*	0.900	0.631
Head length	0.910	0.910	0.910	0.910	0.917	0.917	0.917	0.903[l]	0.691	0.691	0.691	0.583
Head breadth	0.908	*0.908*	0.908	0.888	0.880	0.880	0.880	0.877[l]	0.654	0.654	0.654	0.545

[a]Table 25, p. 97; Table 26, p. 98; Table 96, p. 347.

[b]Table p. 347. Newman *et al.* give raw *r*s only for separated MZ twins; age is not partialed out, probably because many of the *S*s were adults (mean age = 14.58 ± 1.96 years).

[c]Table 2, p. 146.

[d]Table 1, p. 168.

[e]Table 25, p. 97; Table 26, p. 98; Table 96, p. 347.

[f]The partial *r*s were not given by Newman *et al.* (see footnote *b*). The partial *r*s given here were computed by Jensen from the data given by Newman *et al* in Table 87, between pp. 334 and 335, using the formula on p. 95.

[g]This correlation of 0.767 is the Newman *et al.* $r = 0.670$ corrected for restriction of range by McNemar (1938), because the SD = 13.0 of Newman's MZ twins reared apart is so much smaller than the SD for Newman's MZ twins reared together (SD = 17.3) or DZ twins reared together (SD = 15.7), or the population $\sigma = 16$. The logic of the correction for restriction of range is as follows: $r = 1 - (s_w^2/s_t^2)$, where r is the obtained twin correlation, s_w^2 is the variance within twin pairs, and s_t^2 is the total variance of the twin sample. To correct the r for restriction of range (i.e., to show that it would be if the total variance were different from the actual sample variance), one substitutes the new variance (say, the population σ^2) for s_t^2 and solves for r_c (i.e., the corrected value of r).

[h]Stanford Achievement Test.

[i]This $r = 0.583$ is an error. It should be 0.507. The value $r = 0.583$ is the correlation between separated MZ twins on the Woodworth-Mathews personality inventory, which is listed immediately below the Stanford Achievement Test in the Newman *et al.* Table 96, p. 347.

[j]Partial r based on $N = 9$ pairs; height data not given for ten twin pairs.

[k]Partial r based on $N = 11$ pairs; weight data not given for eight twin pairs.

[l]Partial r based on $N = 17$ pairs; head measurements not given for two twin pairs.

by Newman *et al.* (1937). All discrepant correlations in Burt's tabulation are shown here in italics.

As can be seen, at one time (1955) Burt listed the raw r and at another time (1966) the partial r given by Newman *et al.*, but this is not consistent for the different types of twins. One of the rs (Stanford-Binet IQ for MZ Apart, 1966, $r = 0.767$) is a correction of the original correlation for restriction of range. Burt adopted this correction for restriction of range from McNemar (1938), who recommended it because the standard deviation of the Newman *et al.* twins is only 13.0 as compared with a SD of 14.8 for Burt's twin sample and of $\sigma = 16$ for the general population (see footnote *g* in Table II). There seems to be no apparent systematic basis for Burt's selection of the Newman *et al.* rs for comparison with his own rs. The reason for the seemingly unsystematic inconsistencies revealed in Table II can now only be a matter for idle speculation.

Monozygotic Twins Reared Apart

Correlations for monozygotic twins reared apart are shown in Table III. A striking feature in this table is the invariance of the $r = 0.77$, despite the changing sample size.[4] There are three distinct rs, each with a different N (presumably cumulative from one study to the next); the first $r = 0.77$ (based on the Binet test) and the later two (based on a group test) are both 0.771. (The 1956 and 1958 rs are clearly just a repetition of the 1955 r.) It seems improbable, to be sure; but this is hard to judge, for two reasons: (1) the rs are not a random sample of all possible rs, but a sample of rs where the population (or "true") value ρ is probably very close to 0.77, and in this range of high correlations the sampling error is much smaller than for low values of ρ, and (2) the Ns going from 1943 to 1955 to 1966 are cumulative, so that the added cases are much less likely to result in variations in the newly computed rs than if the rs were based on completely independent samples drawn from the same population.

It would seem even more improbable that a totally independent sample of monozygotic twins reared apart would show an intelligence test r of 0.77, and yet this is precisely the r reported by Shields (1962) in his study of 38 pairs of monozygotic twins reared apart, a study that had no connection whatever with Burt's. The correlation between monozygotic twins reared apart may be regarded as an estimate of the broad heritability, h^2 (i.e., the proportion of total phenotypic variance attributable to genotypic variation). It is therefore

[4] I believe the first person to point out this invariance of $r = 0.77$ across changing Ns was Dr. Leon Kamin, in a colloquium of the Psychology Department at the University of Pennsylvania, on September 19, 1972.

Table III. Monozygotic Twins Reared Apart

	1943, p. 91		1955, p. 168 Table 1		1956, p. 125		1958, p. 6 Table 1		1966, p. 146 Table 2		1971, p. 15 Table 4		1972, p. 186 Table 3	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*[a]	*N*	*r*	*N*	*r*
Intelligence														
Group test			21	0.771	?	0.7706	?	0.771	53	0.771				
Individual test	15	0.77	21	0.843			?	0.843	53	0.863				
Final assessment			21	0.876	?	0.8756	?	0.876	53	0.874[b]	?	0.874	53	0.874
Scholastic														
Reading and spelling			21	0.647			?	0.647	53	0.597				
Arithmetic			21	0.723			?	0.723	53	0.705				
General attainments			21	0.681			?	0.681	53	0.623	?	0.623	53	0.623
Physical														
Height			21	0.951					53	0.943	?	0.943	53	0.943
Weight			21	0.897					53	0.884				
Head length			?[c]	0.959					53	0.958				
Head breadth			?[c]	0.962					53	0.960				
Eye color			?[c]	1.000					53	1.000				

[a]Figures for boys and girls have been calculated separately and then averaged.
[b]This is apparently a double-entry Pearson r, which, computed from Burt's data in Table 4, comes out exactly 0.8749 by my own calculation. The unbiased intraclass $r = 0.8771$.
[c]1955, p. 167, footnote 3 states: "The figures for head-length, head-breadth, and eye color are based on much smaller numbers in every batch."

interesting that the estimate of h^2 for a group intelligence test (Otis) arrived at by Newman *et al.* on the basis of monozygotic and dizygotic twin correlations is precisely 0.77 (Newman *et al.*, 1937, Table 39, p. 119). And when the Newman *et al.* correlation between monozygotic twins reared apart is corrected for restriction of range, as was suggested it should be by McNemar (1938), it, too, is 0.77.

Burt's correlations for the final assessments also changed very little from 1955 (with $N = 21$) to 1966 (with $N = 53$). From the final assessments for each of the 53 sets of twins provided by Burt (see Table IV), his 1966 $r = 0.874$ is correct within normal rounding error. (I get $r = 0.8749$ using Pearson's double-entry method for computing r. The unbiased intraclass $r = 0.8771$.)

Burt never published the data on the individual sets of twins, but he sent a list of the final assessments, along with the social class ratings, to those who requested them.[5] This list, in the form prepared by Burt, is presented in Table IV. Actually, not all of the twins listed under "Own home" were reared in their own homes (i.e., by their biological parents, or parent), although the majority were. A few of the twins in this column were reared by a relative who adopted them or by what Burt referred to as "the better type of foster-parent" (Burt, 1966, p. 143). Burt also noted that "we included in our group of separated twins no cases in which both had been brought up by a relative, except for five in which one relative lived in a town and the other in the country" (p. 143). The social classification was based on the occupational categories of the children's parents or foster parents. There is an apparent error in Burt's Table 1 (1966, p. 143), showing the correlational scatter-diagram table for the occupational levels of the twins reared by "own parents" (on the ordinate) and of the twins reared by "foster parents" (on the abscissa); it does not agree completely with the social class ratings presented in Table IV here.

The correlation (double-entry Pearson r) between the social class rating of the twins' homes is -0.084 (omitting cases in residential institutions). The correlation between social class rating and IQ (i.e., final assessments) is -0.314 for twins reared by their "own parents" and $+0.009$ for twins reared in foster homes (omitting those in residential institutions).

[5]The first person to request and receive this list from Burt, as best as I can determine, was Professor William Shockley. Burt later sent me the list in the course of our correspondence about one of my articles on twins (Jensen, 1970; 1972*b*, pp. 307-326). Burt also sent these data to Jencks *et al.* (1972, p. 317, footnote 10). In my article, which made use of Burt's data, I rearranged the order of the twins in each pair so as to make the means and standard deviations of the two arrays (X and Y) as nearly alike as possible, in order that my graphic presentation of the data in the form of a bivariate scatter diagram of the 53 pairs would be as symmetrical as possible, thus better reflecting the properties of the double-entry or intraclass correlation which was used as the measure of relationship. In a later discussion with Burt, he agreed this was probably the best way to make a graphic representation of the twin correlation.

Table IV. Burt's Final Assessments (on IQ Scale) of Intelligence of Monozygotic Twins Reared Apart

	Reared in					Reared in			
	Own home		Foster home			Own home		Foster home	
Case No.	"IQ"	Social class	"IQ"	Social class[a]	Case No.	"IQ"	Social class	"IQ"	Social class
1	68	6	63	6	28	97	5	92	1
2	71	4	76	5	29	97	3	95	5
3	73	5	77	5	30	97	5	112	R
4	75	6	72	R[b]	31	97	6	113	5
5	78	3	71	6	32	99	5	105	4
6	79	3	75	5	33	100	3	88	3
7	81	5	86	5	34	101	5	115	5
8	82	2	82	R	35	102	5	104	R
9	82	4	93	2	36	103	6	106	5
10	83	4	86	6	37	105	6	109	1
11	85	5	83	6	38	106	5	107	5
12	86	6	94	6	39	106	6	108	R
13	87	5	93	1	40	107	5	108	3
14	87	6	97	6	41	107	3	101	2
15	89	6	102	2	42	108	5	95	5
16	90	2	80	2	43	111	6	98	5
17	91	3	82	2	44	112	6	116	5
18	91	2	88	5	45	114	1	104	5
19	92	5	91	6	46	114	5	125	5
20	92	2	96	3	47	115	2	108	6
21	93	6	87	2	48	116	2	116	6
22	93	3	99	4	49	118	2	116	6
23	93	5	99	3	50	121	1	118	5
24	94	6	94	R	51	125	4	128	5
25	95	6	96	5	52	129	2	117	5
26	96	2	93	R	53	131	1	132	4
27	96	4	109	1					

[a]Social class 1 = high, 6 = low.

[b]R denotes residential institution.

One may wonder to what extent Burt's final assessments might differ from IQs obtained from a single administration of a standard intelligence test. Unfortunately, we do not have the actual individual test scores on Burt's twins (these were reported for only one pair: Burt, 1966, p. 144), so that we could determine the correlation between the single test scores and the final assessments.

The most crucial point in twin studies, of course, is the difference in scores between the members of each twin pair, and we can look to see if these twin differences for Burt's final assessments are in any way atypical of the twin differences in IQ found by other investigators who used only actual test scores rather than any kind of adjusted assessment. There are only three studies besides

Burt's in the literature of monozygotic twins reared apart, totaling 69 pairs, which we can use as the basis for comparison with Burt's final assessments.[6] Figure 1 shows the frequency distribution of the absolute IQ differences between twins in the three studies, along with those of Burt. The question is, do Burt's data differ significantly from the data of the three other studies? An analysis of variance shows that the mean twin differences of the four studies do not differ significantly ($F = 0.8732$, df = 3/118, $p = 0.4572$). To test if moments other than the mean differ between Burt's and the others' data, the Kolmogorov-Smirnov test is the most appropriate. This is a nonparametric test of whether two independent samples can be regarded as drawn from the same population or from populations with the same distribution; it is simultaneously sensitive to any kind of differences in the distributions–central tendency, dispersion, skewness, etc. When the distribution of Burt's twin differences is compared with the distribution of the twin differences in the three other studies combined, the Kilmogorov-Smirnov test yields a χ^2 value of 1.50 (df = 2), which is statistically nonsignificant. (With 2 df, a χ^2 value of 5.99 or greater is required for significance at the 0.05 level.) Therefore, it may be concluded that the distribution of twin differences in Burt's sample does not differ significantly from the distribution of twin differences in the three other studies.

Dizygotic Twins Reared Together

The puzzling point in Table V is the $N = 172$ in 1955 and $N = 127$ in 1966. What became of the other 45 pairs of dizygotic twins? Or is this a misprint, the *127* being merely a reversal of *172*? Since the *r*s differ from 1955 to 1966, presumably somewhat different samples are involved. Also, the reprint of the 1966 article given to me by Burt has a meticulously inscribed correction of one of the *r*s in Table 2 (p. 146) (see footnote *c* in Table V), and I note that the same correction has been made in a reprinted version of the 1966 article (in Manosevitz *et al.*, 1969, p. 333). If $N = 127$ had been an error, then it would presumably have been corrected as well. So it seems more likely that the 1955 $N = 172$ is the more questionable. Also, note that the correlations for arithmetic, general attainments, height, and weight are invariant from 1955 to 1966, despite the reduction of N from 172 to 127. This would seem to make it extremely likely that the 1955 $N = 172$ is in error and should be $N = 127$. If that is the case, however, why do four of the correlations not remain invariant from 1955 to 1966? Burt's writings contain no clue to the answer.

Burt's 1943 $N = 156$ makes one wonder if perhaps cases were lost or for some reason discarded between 1956 and 1966. The fact that Burt (1966) notes

[6] I have presented elsewhere detailed analyses of the original data from all four major studies of monozygotic twins reared apart (Jensen, 1970; 1972*b*, pp. 307-326).

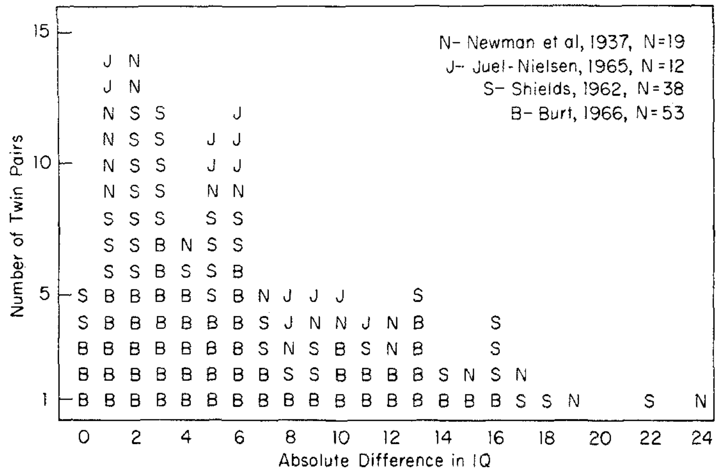

Fig. 1. Distribution of absolute differences in IQ between 122 pairs of co-twins reared apart, in four studies. Twin pairs in the various studies are indicated by the author's initial.

that N = 71 for the same-sex twins and N = 56 for different-sex twins is consistent with the total N = 127.

Full Siblings Reared Together

Table VI presents the same puzzle, i.e., a decrease in N from the 1955 N = 1000 and the 1956 N = 987 to the 1966 N = 264. The decrease from 1966 to 1972 (Table 1) is less puzzling, since Burt mentions (1972, p. 187) that the r of Table 5 is based on a larger and more recent sample than the r of Table 1 (which is the same in the 1966 data). But it seems puzzling that Burt should report correlations on some 800-1000 siblings in 1955 and 1956 and then report rs in 1966 and 1972 based on Ns of only 264 and 231. I wish I had noticed this peculiarity while Burt was alive so that I could have asked for an explanation; I have searched his writings in vain for one. The invariance for final assessment r = 0.507 from the 1955 N = 1000 to the 1956 N = 987 leaves the question of why three pairs of sibs were dropped; but the invariance of r is not surprising, since with N = 1000, the loss of three cases, unless they were extremely atypical, would not be likely to change the r to the third decimal. The r = 0.568 for weight is also invariant from the 1955 N = 853 to the 1966 N = 264, which seems improbable, though of course entirely possible, especially when the rs vary only slightly around some central tendency, i.e., the population value, as seems to be the case in these data. A possible explanation for these discrepancies in Ns is that in different studies the reported N might be (*1*) the number of *individuals* in the sample, or (*2*) the number of pairs of siblings entering into the correlation, or (*3*) the number of families involved.

Table V. Dizygotic Twins Reared Together

	1943, p. 91		1955, p. 168 Table 1		1956, p. 124		1958, p. 6 Table 1		1966, p. 146 Table 2	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	r^a
Intelligence										
Group test			172	0.542	?	0.542	?	0.542	127[b]	0.552
Individual test	156	0.54	172	0.526			?	0.526	127	0.527
Final assessment			172	0.551	?	0.551	?	0.551	127	0.453[c]
Scholastic										
Reading and spelling			172	0.915			?	0.915	127	0.919
Arithmetic			172	0.748			?	0.748	127	0.748
General attainments			172	0.831			?	0.831	127	0.831
Physical										
Height			172	0.472					127	0.472
Weight			172	0.586					127	0.586
Head length			?[d]	0.495					100	0.495
Head breadth			?[d]	0.541					100	0.541
Eye color			?[d]	0.516					100	0.516

[a]Figures for boys and girls have been calculated separately and then averaged.

[b]1966, Table 4, p. 150, *states*: "Same sex $N = 71$, different sex $N = 56$."

[c]This is a misprint, later corrected by Burt; correct $r = 0.534$. (This correction also occurs in reprinted version of 1966 in Manosevitz *et al.*, 1969, p. 333).

[d]1955, p. 167, footnote 3 states: "The figures for head-length, head-breadth, and eye color are based on much smaller numbers in every batch."

Table VI. Full Siblings Reared Together

	1955, p. 168 Table 1		1955, p. 173		1956, p. 119 Table 10		1958, p. 6 Table 1		1966, p. 146 Table 2		1966, p. 150 Table 4		1972, p. 182 Table 1		1972, p. 187 Table 5	
	N	*r*	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*[a]	*N*	*r*	*N*	*r*	*N*	*r*
Intelligence																
Group test	853	0.515					?	0.515	264	0.545						
Individual test	853	0.491					?	0.491	264	0.498						
Final assessment	853	0.538	1000[b]	0.507	987	0.507[c]	?	0.538	264	0.531						
Unspecified											264	0.53	231	0.543	?	0.531[d]
Scholastic																
Reading and spelling	853	0.853					?	0.853	264	0.842						
Arithmetic	853	0.769					?	0.769	264	0.754						
General attainments	853	0.814					?	0.814	264	0.803			231	0.685	?	0.803[e]
Physical																
Height	853	0.503							264	0.501			231	0.483	?	0.501[f]
Weight	853	0.568							264	0.568						
Head length	?[g]	0.481							100	0.481						
Head breadth	?[g]	0.507							100	0.510						
Eye color	?[g]	0.553							100	0.554						

[a]Figures for boys and girls have been calculated separately and then averaged.
[b]Limited to children aged 8-13.
[c]Given as 0.5069 on p. 125. The r for unadjusted scores (usually meaning the group test) is given as 0.5342 (p. 125).
[d]1972, p. 189, states this r is from Burt (1966) (final assessment $r = 0.531, N = 264$).
[e]1972, p. 187, states this r is from Burt (1966) (general attainments $r = 0.803, N = 264$).
[f]1972, p. 187, states this r is from Burt (1966) (height = 0.501, $N = 264$).
[g]1955, p. 167, footnote 3 states: "The figures for head-length, head-breadth, and eye color are based on much smaller numbers in every batch."

Full Siblings Reared Apart

Table VII shows an increased N but invariant rs for the scholastic measures in 1955 and 1966. One of the Ns (most probably 1966) for the scholastic and physical measures is certainly in error, since all five rs are invariant from the 1955 N = 131 to the 1966 N = 151. The true N for 1966, at least for the scholastic and physical measures (height and weight only) is surely better taken to be 131 instead of 151. Since Burt carefully noted a correction in the reprint of the 1966 table he sent to me (see present Table V, footnote *c*), why was not N = 151 also corrected if it was in error? Probably the N = 151 applies only to the intelligence measures and not to the scholastic and physical measures, for which the true N is most probably 131.

Another point of question in Table VII is the 1972 N = 156. Is this just an addition of five cases to the 1966 N = 151? It is doubtful that a 3% increase in N would alter the r by more than 0.10, as is here the case. If the true 1966 N were 131, however, and there were an addition of 25 cases (a 20% increase), the changed r would not be so improbable.

In Burt's Table 4 (1966, p. 150), under "other investigators," Burt gives the "number of investigations" of siblings reared apart as 33; I questioned this to Burt and he said it was a misprint—the correct number is 3.

Unrelated Children Reared Together

Table VIII also presents some difficulties. The 1955 N is not clear, as indicated in footnote *b*. Since two of the rs are invariant from 1955 to 1966, one probably should assume that the 1955 N = 136. The difference in correlations for the final assessments (0.269 *vs.* 0.267) is small enough to be within the range of differences that can result from using different methods for computing r or by computing r for the combined sexes, in one case, and computing r for the sexes separately and averaging them, in the other. The invariance or very close similarity in all the rs for scholastic and physical measures reinforces the high probability that the 1955 N = 136, although there is nothing in the 1955 article that gives this information; the only N mentioned in the text is 287, but it is not explicit that all of the 287 foster children on whom data were secured are represented in the correlations presented in Burt's Table 1.

Miscellaneous Kinship Correlations

Some of the most interesting and more rarely found kinship correlations, from a genetic standpoint, are shown in Table IX. Seldom did Burt specify the particular intelligence measures used in these correlations. I have indicated the tests in footnotes when they were explicitly indicated by Burt. We again see the

Table VII. Full Siblings Reared Apart

	1955, p. 168 Table 1		1958, p. 6 Table 1		1966, p. 146 Table 2		1966, p. 150 Table 4		1972, p. 182 Table 1	
	N	*r*	*N*	*r*	*N*	*r*[a]	*N*	*r*	*N*	*r*
Intelligence										
Group test	131	0.441	?	0.441	151	0.412				
Individual test	131	0.463	?	0.463	151	0.423				
Final assessment	131	0.517	?	0.517	151	0.438				
Unspecified							151	0.44	156	0.468
Scholastic										
Reading and spelling	131	0.490	?	0.490	151	0.490				
Arithmetic	131	0.563	?	0.563	151	0.563				
General attainments	131	0.526	?	0.526	151	0.526			156	0.401
Physical										
Height	131	0.536			151	0.536			156	0.431
Weight	131	0.427			151	0.427				
Head length	?[b]	0.536			100	0.506				
Head breadth	?[b]	0.472			100	0.492				
Eye color	?[b]	0.504			100	0.524				

[a]Figures for boys and girls have been calculated separately and then averaged.

[b]1955, p. 167, footnote 3 states: "The figures for head-length, head-breadth, and eye color are based on much smaller numbers in every batch."

Table VIII. Unrelated Children Reared Together

	1955, p. 168 Table 1		1958, p. 6 Table 1		1966, p. 146 Table 2		1966, p. 150 Table 4	
	N	r	N	r	N	r^a	N	r
Intelligence								
Group test	287[b]	0.281	?	0.281	136	0.281		
Individual test	287[b]	0.252	?	0.252	136	0.252		
Final assessment	287[b]	0.269	?	0.269	136	0.267		
Unspecified							136	0.27
Scholastic								
Reading and spelling	287[b]	0.548	?	0.548	136	0.545		
Arithmetic	287[b]	0.476	?	0.476	136	0.478		
General attainments	287[b]	0.535	?	0.535	136	0.537		
Physical								
Height	287[b]	-0.069			136	-0.069		
Weight	287[b]	0.243			136	0.243		
Head length	?[c]	0.116			100	0.110		
Head breadth	?[c]	0.082			100	0.082		
Eye color	?[c]	0.104			100	0.104		

[a]Figures for boys and girls have been calculated separately and then averaged.

[b]This N is rather ambiguously reported: "she [Miss Conway] also secured data for 287 foster children" (1955, p. 167). Is this then the N for the correlation between "unrelated children reared together" given in Table 1, p. 168? Or are "unrelated children reared together" only a subset of the total number of foster children?

[c]1955, p. 167, footnote 3 states: "The figures for head-length, head-breadth, and eye color are based on much smaller numbers in every batch."

enigmatic shrinking Ns between 1955 or 1956 and 1966 for some of the kinships.

The 1955 parent-offspring $r = 0.481$ with $N = 954$ is based on test scores for the children and *interview assessments* of the intelligence of at least one parent. Burt (1955, p. 172, footnote 1) states: "For the assessments of the parents we relied chiefly on personal interviews; but in doubtful or borderline cases [no N indicated] an open or a camouflaged test was employed." The rs for reduced Ns reported subsequently to 1955 might conceivably be for parent-child pairs in which both parents and children had been tested, but this would seem an unreasonable kind of sampling if the tested parents were only those who were "doubtful or borderline cases."

Was the interview assessment of intelligence (of which the reliability and validity are not specifically reported) also used in other cross-generational correlations, e.g., uncle-nephew and grandparent-grandchild?

Table IX. Various Kinship Correlations for Intelligence Measures and Physical Stature

	Intelligence measures												Stature			
	1943, p. 91		1955, p. 173		1956, p. 119 Table 10		1966, p. 150 Table 4		1971, p. 15 Table 4		1972, p. 186 Table 3		1956, p. 117 Table 9		1972, p. 186 Table 3	
Relationship	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*[a]	*N*	*r*	*N*	*r*	*N*	*r*	*N*	*r*
Parent (as adults) and offspring					963	0.489[b]	374	0.49[c]								
Parent (as children) and offspring							106	0.56[c]								
Parent (age unspecified) and offspring			954[d]	0.481					?	0.492[e]	374	0.492[f]	?	0.507	374	0.507
Grandparent and grandchild					321	0.335[c]	132	0.33[c]					263	0.322		
Dizygotic twins–same sex							71	0.55[c]								
Dizygotic twins–different sex							56	0.52[c]								
Uncle Niece Aunt Nephew					375	0.354[c]	161	0.34[c]					615	0.287		
First cousins	167	0.30[g]			552	0.289[c]	215	0.28[c]					584	0.236		
Second Cousins	86	0.24[g]					127	0.16[c]								
Fathers and mothers (assortative mating coefficient)					?	0.4526[h]			?	0.379[c,i]	95	0.379[j]			95	0.280
Siblings born to first cousins					287	0.623[c,k]										
Brothers and sisters in same foster homes	157/?[l]	0.51														
Brothers and sisters in different foster homes	157/?[l]	0.42														
Foster children with foster parent's own children[d]	157/?[l]	0.27														
Foster parent and foster child							88	0.19[c]								
Unrelated children reared apart							200	-0.4[c]								

[a]Figures for boys and girls have been calculated separately and then averaged.
[b]"Adjusted assessment," also given in text (p. 125) as $r = 0.4887$. Unadjusted $r = 0.5288$ (p. 125).
[c]Intelligence test unspecified.
[d]Limited to children between ages 8 and 13. Verbal and nonverbal tests of intelligence.
[e]"Adjusted assessment." For "attainments," $r = 0.514$.
[f]For "attainments," $N = 374$, $r = 0.514$.
[g]Binet test.
[h]1956, p. 125, r for unadjusted scores. Adjusted assessments $r = 0.386$ (p. 118); also given as $r = 0.3875$ for adjusted assessments (p. 125).
[i]For "attainments," $r = 0.678$.
[j]For "attainments," $N = 95$, $r = 0.678$.
[k]As compared with $r = 0.507$ for ordinary siblings (1956, p. 124).
[l]$N < 157$, since the total N of all foster children equals 157.

Is the 1956 $N = 321$ (grandparent and grandchild) transposed as the 1966 $N = 132$ for which the r is invariant to two places?

The correlation (0.623, $N = 287$) between siblings resulting from first-cousin matings, as far as I can determine, is unique in the genetic literature. (This odd kinship correlation is not reported but is latent in the massive data secured by Schull and Neel, 1965, on the offsprings of cousin marriages in Japan.)

CONCLUSION

Unfortunately, since Burt is deceased, it seems highly unlikely that we shall ever be able to clear up the rather puzzling discrepancies and ambiguities that were noted in the above tables. Nor shall we be able to determine more exactly the relationship between the "final assessments" and the actual test scores on which they were based. The only "scores" on individual sets of relatives that remain are the final assessments of the monozygotic twins reared apart, which Burt made available a few years before his death.

But the most serious problems with Burt's presentation of all these correlations are the often unknown, ambiguous, or inconsistent sample sizes and the invariant correlations despite varying Ns from one report to another. I count altogether no fewer than 20 pairs of invariant correlations for various kinships with differing Ns in each case. If the Ns are questionable, the standard errors of the correlations are necessarily in doubt, and without estimates of the standard error, *ipso facto*, the correlations are useless for hypothesis testing. Unless new evidence rectifying the inconsistencies in Burt's data is turned up, which seems doubtful at this stage, I see no justifiable alternative conclusion in regard to many of these correlations. Hypothesis testing depends on data of determinate reliability. Of this, Burt's presentation of his own data unfortunately often gives too little assurance. Any particular instance of an invariant r despite a changed N can be rationalized as being not too improbable. But 20 such instances unduly strain the laws of chance and can only mean error, at least in some of the cases. But error there surely must be. Usually, we have no sure way to distinguish the erroneous from the valid Ns. Be conservative and pick whichever N is smaller? Perhaps even more disturbing is the doubt that so many questionable and discrepant points cast on the trustworthiness of the various correlation coefficients themselves.

In hopes that some of the original data might be recovered after Burt's quite sudden death in October 1971, I corresponded with Burt's personal secretary about this possibility and also looked into it further when I was in London the following summer. But by then, alas, nothing remained of Burt's possessions save various notes, letters, manuscripts, reprints, and books. I was told that shortly after Burt's death many of the books and journals had to be sold and donated to university libraries, and that many boxes of old data, which Burt had kept for many years, were disposed of in the course of vacating his flat in Hampstead.

These boxes, etc., I was informed, were either poorly labeled or not labeled at all, so that their exact contents were not apparent to casual inspection. And so, unfortunately, the original data are lost, and all that remains are the results of the statistical analyses, which are presented in the foregoing tables as completely, I believe, as possible.

An item of information that is lacking and would be of value for certain genetic analyses is the variances (or standard deviations) of the test scores for the various kinships; with these, along with the correlations, one could compute the regressions, which are more useful than correlations in certain analyses, for making correction for restriction of range, and so forth.

The reporting of kinship correlations at times with and at times without noting the sample size, the rather inconsistent reporting of sample sizes, the higher than ordinary rate of misprints in Burt's published tables (several of them acknowledged by himself), and the quite casual description of the tests and the exact procedures and methods of data analysis all stand in quite strange and marked contrast to the theoretical aspects of Burt's writings in this field, which were elegantly and meticulously composed, with profound erudition and impressive technical sophistication. It is almost as if Burt regarded the actual data as merely an incidental backdrop for the illustration of the theoretical issues in quantitative genetics, which, to him, seemed always to hold the center of the stage.

In the theoretical aspects of the applications of quantitative genetics to psychometric data, Burt was outstandingly ahead of all others of his time, as he was, too, in psychometric theory in general. To read his major articles is indeed an education in itself. He was undoubtedly the first psychologist to understand thoroughly, and to use, the important contributions of Fisher, Haldane, and Mather in biometrical genetics, a complex and difficult science which Burt so ably explicated in several of his publications, most notably in the extensive 1956 article with Margaret Howard and in one of his last major articles (Burt, 1971). In behavioral genetics, Burt's classical theoretical approach has been technically superseded only by the recent applications of biometrical genetics to human behavioral data by the Birmingham group (e.g., Eaves, 1969; Jinks and Fulker, 1970; Mather and Jinks, 1971).

The Newman *et al.* (1937)[7] monograph on twins is an interesting contrast to Burt's works. Newman *et al.* reported all of the basic scores on their twins and

[7]Newman was principally an embryologist, Freeman a psychologist, and Holzinger a statistician. It was Holzinger who was responsible for all the quantitative treatment of the data. He went about this like a good statistician, but without showing the slightest evidence that he was aware of anything in the field of quantitative genetics, which at that time was already quite advanced and would have been highly applicable to the excellent data gathered by Newman *et al.* It should be noted, too, that this study was conducted at the University of Chicago at the time that Sewall Wright, a leader in quantitative genetics, was professor of genetics there; but apparently Wright was not consulted by Newman *et al.* This is one of those interesting curiosities in the history of science.

were thoroughly explicit in all matters of test description, data collection procedures, and statistical analyses; but the analyses are all so simple and naive from a genetic standpoint that a reader of Newman *et al.* would never guess the existence of a science of quantitative genetics. Not a single reference is made, for example, even to the important theoretical and methodological contributions of Sewall Wright and R. A. Fisher, particularly the latter's seminal 1918 paper, "The Correlations Between Relatives on the Supposition of Mendelian Inheritance," which Burt made so much use of and which has since come to be generally regarded as one of the cornerstones of quantitative genetics. Newman *et al.* seemed unaware of such genetically important considerations and complications as assortative mating, dominance deviation, and genotype X environment interaction and covariance. Moreover, thanks to Holzinger's simple and often cited, but theoretically nonsensical, heritability index, all the heritabilities reported in Newman *et al.* purely and simply are wrong; i.e., they do not represent estimates of what any geneticist has ever meant by "heritability," *viz.*, the ratio of genotypic variance to phenotypic variance. (For a formal proof, see Jensen, 1972*b*, pp. 294-306, especially the footnote on p. 295.)

But there is at least an important lesson in all of this for present and future researchers. For while Newman *et al.*, even in 1937, were behind the times in the theoretical treatment of their twin data, their data at least were sufficiently explicit and fully preserved so that other researchers could use them with confidence in theoretically more satisfactory ways. On the other hand, Burt, who collected much more data and a wider variety of data of considerably greater theoretical value, and treated them in terms of the currently most sophisticated genetic theory, in a sense may have performed less of a service strictly to empirical science than did Newman *et al.*, by not taking pains to report his results in a more complete and accurate form and to insure the preservation of the original data for the benefit of later investigators.

Quantitative genetics provides the basic methodology for behavioral genetics. As the theoretical concepts of quantitative genetics advance and change, new or modified analytical methods can be applied to fundamental problems, and old data collected in the past and treated with less adequate methods can be reanalyzed with new and improved methods, or can be combined with data from other sources in order to increase the statistical reliability of the results. For this reason, it is important that investigators preserve for future use the kinds of kinship data that are often so difficult and costly to obtain.

Especially rare data, such as those of monozygotic twins reared apart, siblings from cousin matings, double first cousins, and the offspring of two mated pairs of monozygotic twins (genetically, such offpsring are full siblings with entirely different parents), should be published in full, along with complete descriptions of the tests or measurements and procedures. Perhaps this should be a general requirement for the publication of studies based on such valuable data, so that quantitative analytical techniques other than those used by the original author can be applied to the data by anyone who wishes. Less rare kinship data, such as ordinary monozygotic and dizygotic twins, siblings, parent-child, cousins, etc., should probably be submitted complete in some standard form to the journals publishing the studies based on these data, to be permanently preserved in the journals' archives for future workers who may wish to use them. They should be accompanied by specimen sets of the measuring instruments and descriptions of the procedures used to obtain the data, along with a full description of the population sampled and the sampling technique employed. In the case of twins studies, the methods and data for determining zygosity should be given.

Ancillary information on every subject, such as age and sex and race, would, of course, be essential; and other information, such as date of testing, socioeconomic background, educational level, and geographic location also would be desirable, especially as a basis for decisions about combining data from various sources.

It would also be wise in all kinship studies routinely to obtain standardized measurements on each subject of at least one or two highly heritable metric physical characteristics, such as height, head length and breadth, and fingerprint ridges, to which one could apply the same analytical methods as were applied to the behavioral traits in question. As originally suggested by R. A. Fisher, such a parallel analysis of physical traits that are little influenced by the environment may serve as a sheet anchor for the quantitative genetic interpretation of the less heritable behavioral traits under consideration. If the particular genetic analysis does not make sense for the highly stable physical traits, it is probably also questionable for the less stable behavioral traits, either because the sampling has been faulty or because the particular analytical model is inappropriate.

Observance of these recommendations should aid the advancement of behavioral genetics.

REFERENCES

Bulmer, M. G. (1970). *The Biology of Twinning,* Clarendon Press, Oxford.

Burt, C. (1921). *Mental and Scholastic Tests,* P. S. King, London (4th ed. 1962, Staples Press).

Burt, C. (1923). *Handbook of Tests for Use in Schools,* King & Son, London (2nd ed. 1948, Staples Press).

Burt, C. (1943). Ability and income. *Brit. J. Educ. Psychol.* 13: 83-98.

Burt, C. (1955). The evidence for the concept of intelligence. *Brit. J. Educ. Psychol.* 25: 158-177.

Burt, C. (1958). The inheritance of mental ability. *Am. Psychologist,* 13: 1-15.

Burt, C. (1966). The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *Brit. J. Psychol.* 57: 137-53.

Burt, C. (1971). Quantitative genetics in psychology. *Brit. J. Math. Stat. Psychol.* 24: 1-21.

Burt, C. (1972). Inheritance of general intelligence. *Am. Psychologist* 27: 175-190.

Burt, C., and Howard, M. (1956). The multifactorial theory of inheritance and its application to intelligence. *Brit. J. Stat. Psychol.* 9: 95-131.

Burt, C., and Howard, M. (1957). The relative influence of heredity and environment on assessments of intelligence. *Brit. J. Stat. Psychol.* **10**: 99-104.

Eaves, L. J. (1969). The genetic analysis of continuous variation: A comparison of experimental designs applicable to human data. *Brit. J. Math. Stat. Psychol.* **22**: 131-147.

Fisher, R. A. (1934). *Statistical Methods for Research Workers,* Oliver & Boyd, Edinburgh.

Haggard, E. A. (1958). *Intraclass Correlation and the Analysis of Variance,* Dryden Press, New York.

Honzik, M. P. (1957). Developmental studies of parent-child resemblance in intelligence. *Child Develop.* **28**: 215-228.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., and Michelson, S. (1972). *Inequality: A Reassessment of the Effect of Family and Schooling in America,* Basic Books, New York.

Jensen, A. R. (1968). Social class, race, and genetics: Implications for education. *Am. Educ. Res. J.* **5**: 1-42.

Jensen, A. R. (1970). IQs of identical twins reared apart. *Behav. Genet.* **1**: 133-148.

Jensen, A. R. (1972*a*). Sir Cyril Burt (obituary). *Psychometrika* **37**: 115-117.

Jensen, A. R. (1972*b*). *Genetics and Education,* Methuen, London.

Jinks, J. L., and Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychol. Bull.* **73**: 311-349.

Juel-Nielsen, N. (1965). Individual and environment: A psychiatric-psychological investigation of monozygous twins reared apart. *Acta Psychiat. Neurol. Scand.,* Monogr. Suppl. 183.

Manosevitz, M., Lindzey, G., and Thiessen, D. D. (1969). *Behavioral Genetics: Method and Research,* Appleton-Century-Crofts, New York.

Mather, K. (1964). *Statistical Analysis in Biology,* 5th ed., Methuen, London.

Mather, K., and Jinks, J. L. (1971). *Biometrical Genetics,* 2nd ed., Chapman & Hall, London.

McNemar, Q. (1938). Special review, Newman, Freeman, and Holzinger's twins: A study of heredity and environment. *Psychol. Bull.* **35**: 237-249.

Newman, H. H., Freeman, F. N., and Holzinger, K. J. (1937). *Twins: A Study of Heredity and Environment,* University of Chicago Press, Chicago.

Nichols, R. C., and Bilbro, W. C., Jr. (1966). The diagnosis of twin zygosity. *Acta Genet.* **16**: 265-275.

Pearson, K. (1901). Mathematical contributions to the theory of evolution. IX: On the principle of homotyposis and its relation to heredity, to variability of the individual, and to that of race. Part I: Homotyposis in the vegetable kingdom. *Phil. Trans. Roy. Soc. (Lond.) Ser. A* **197**: 285-379.

Schull, W. J., and Neel, J. V. (1965). *The Effects of Inbreeding on Japanese Children,* Harper & Row, New York.

Shields, J. (1962). *Monozygotic Twins Brought Up Apart and Brought Up Together,* Oxford University Press, London.