29 September 1978, Volume 201, Number 4362 SCIENCE

The Cyril Burt Question: New Findings

The eminent Briton is shown, beyond reasonable doubt, to have fabricated data on IQ and social class.

D. D. Dorfman

Cyril Burt had a powerful impact on the history of quantitative and empirical research on intelligence and its genetic basis. The importance of his work was widely recognized in both Great Britain and the United States, and he was preeminent in his field. In 1946 Burt was knighted for his contributions to education. He was the first psychologist to receive a knighthood.

Summary. Cyril Burt presented data in his classic paper "Intelligence and social mobility" that were in perfect agreement with a genetic theory of IQ and social class. A detailed analysis of these data reveals, beyond reasonable doubt, that they were fabricated from a theoretical normal curve, from a genetic regression equation, and from figures published more than 30 years before Burt completed his surveys.

In a biographical sketch published shortly after his death he was described thus (*1*, p. 117): "The overall picture that Sir Cyril leaves in one's memory, after corresponding with him, seeing him, and conversing with him, is very clear indeed. Everything about the man—his fine, sturdy appearance; his aura of vitality; his urbane manner; his unflagging enthusiasm for research, analysis, and criticism; even such a small detail as his firm, meticulous handwriting; and, of course, especially his notably sharp intellect and vast erudition—all together leave a total impression of immense quality, of a born nobleman."

Recently, Burt has been accused of fabricating some of his most widely cited data (*2*-*5*). First, Kamin (*6*) discovered a number of ambiguities and oddities in Burt's papers, the most remarkable being some kinship correlations that did not change with increasing sample size. A widely quoted example consists of reported correlations on a group intelligence test of monozygotic twins reared apart (*6*, p. 38). In 1955 Burt reported a correlation of .771 for 21 such pairs of twins (*7*). In 1958 he also reported .771 for a sample size "over 30" (*8*, p. 7) and in 1966 again .771 for a sample of 53 pairs (*9*). Fabrication is one explanation of these invariances, but there are other possibilities that are not implausible. For instance, it has been suggested (*10*) that "the most likely hypothesis to account for the flaws in Burt's tables is one mentioned in Wade's article [(*5*)]—being familiar with the literature and realizing that a few erroneous values wouldn't change things much, the aging Burt regrettably took the shortcut of repeating correlations he had calculated earlier rather than taking the trouble to compute new ones." If this view is accepted, Burt was guilty of deception but not of inventing data. Fulker (*11*, p. 511) concluded that "he has probably been guilty more of a careless reporting that stemmed from a disregard for sample size, as Jensen recently concluded, than of any attempt to mislead."

Second, Gillie (*2*, *3*) also imputed fabrication to Burt in suggesting that two of Burt's important collaborators, Jane Conway and Margaret Howard, were fictitious persons. It appears, however, that Howard did exist, and she has been described as "mathematically competent" by John Cohen of the University of Manchester (*12*).

Third, Clarke and Clarke (*13*) and McAskie and Clarke (*14*) have questioned the results on parent-offspring regression reported in Burt's classic paper "Intelligence and social mobility" (*15*). To Clarke and Clarke, the regressions halfway to the mean that Burt reported "appear suspiciously perfect" (*13*, p. 168). McAskie and Clarke noticed that the mean IQ for his higher professional group was 153.2 in 1943 (*16*) and by 1961 "had miraculously shrunk to 139.7, but, surprisingly, the range for the offspring underwent no change during the same period" (*14*, p. 256). In his 1961 paper Burt did not comment on the large drop, which brought the parent-offspring regression for the higher professional group into essentially perfect agreement with his theory. The earlier 153.2 could have been a computational or typographical error. Errors were not rare in Burt's publications (*17*).

Fourth, statistical analyses of data released by Burt on his 53 pairs of monozygotic twins reared apart revealed a negative correlation (−0.084) between the social class ratings of the separated twins' homes (*17*, p. 14). This finding, which argues against possible environmental reasons for the correlation of separated twins' IQ's, is contrary to the usual finding that foster children tend to be placed in homes similar to those of their biological parents (*6*).

The most spirited defense of Burt has been provided by Jensen, who discovered many of the oddities in Burt's data.

The author is professor of psychology in the Department of Psychology, University of Iowa, Iowa City 52242.

0036-8075/78/0929-1177$02.00/0

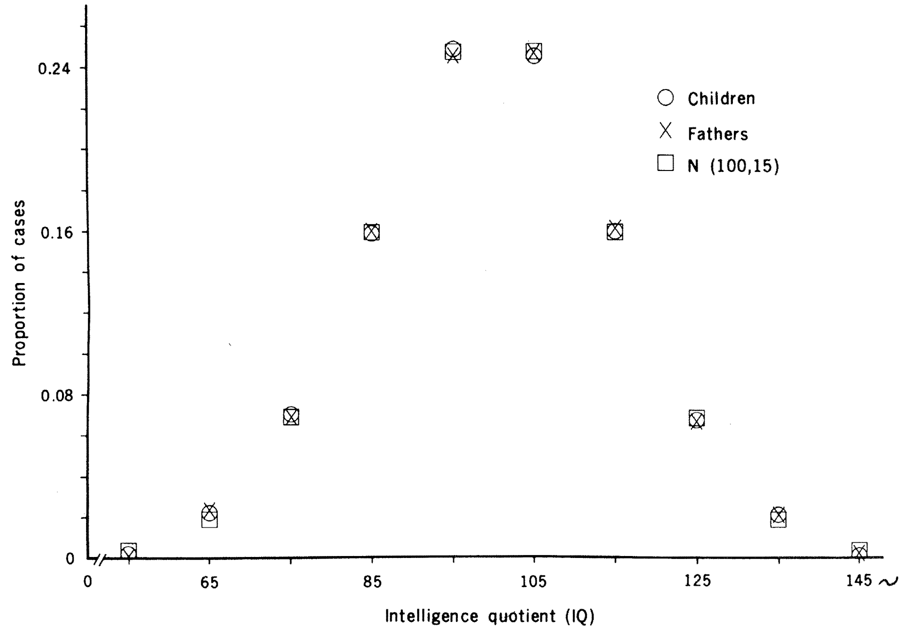

Fig. 1. Burt's frequency distributions of intelligence for the fathers and their children. The proportionate frequencies for the fathers are the column totals of Table 1 divided by 1000 and for the children the column totals of Table 2 divided by 1000. For purposes of comparison, a theoretical normal distribution with mean 100 and standard deviation 15 is also presented.

He declared (*18*, p. 471): "The central fact is that absolutely no evidential support for these trumped-up charges of fakery and dishonesty on the part of Burt has been presented by his accusers. The charges, as they presently stand, must be judged as the sheer surmise and conjecture, and perhaps wishful thinking, of a few intensely ideological psychologists whose antipathy for Burt's hereditarian position in the so-called 'IQ controversy' was already well known to researchers in this field long prior to the *Sunday Times*' broadside. Professor Leon Kamin, who apparently spearheaded the attack, has been trying for several years now, in his many speeches and a book (*The Science and Politics of IQ*), to wholly discredit the large body of research on the genetics of human mental abilities. The desperate scorched-earth style of criticism against genetics that we have come to know in this debate has finally gone the limit, with charges of 'fraud' and 'fakery' now that Burt is no longer here to answer for himself or take warranted legal action against such unfounded defamation." Eysenck has also defended Burt and has said (*19*, p. 675): "A journalist, by issuing an unjustified and completely irresponsible accusation of fraud against a well-known psychologist, has been able to destroy that psychologist's reputation and to throw doubt on well-supported theories through 'guilt by association.' This is the essence of McCarthyite character assassination, and is no more acceptable from the left than it was from the right." Nicholas Wade of *Science* concluded (*5*, p. 919) that "it would still be of some historical interest to know whether the flaws resulted from systematic fraud, mere carelessness, or something in between. The facts so far available do not allow any of these explanations to be ruled out."

In summary, it appears that further evidence is needed if the question is ever to be resolved beyond a reasonable doubt.

Burt's Classic Study

The major goal of this investigation is to analyze in depth the data reported by Burt in "Intelligence and social mobility" (*15*). The data I examine are presented in Burt's tables I to IV, which consist of frequency distributions of intelligence of children and their fathers from six occupational classes. His tables V and VI are simple transformations of the data of tables III and IV used to compute mobility across classes.

The study attracted a great deal of in-

Table 1. Burt's table I (*15*, p. 11), copied exactly except for boldness of type.

TABLE I. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: ADULTS

	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+	Total	Mean I.Q.
I. Higher Professional									2	1	3	139.7
II. Lower Professional							2	13	15	1	31	130.6
III. Clerical				1	8	16	56	38	3		122	115.9
IV. Skilled			2	11	51	101	78	14	1		258	108.2
V. Semiskilled		5	15	31	135	120	17	2			325	97.8
VI. Unskilled	1	18	52	117	53	11	9				261	84.9
Total	1	23	69	160	247	248	162	67	21	2	1000	100.0

Table 2. Burt's table II (*15*, p. 11), copied exactly except for boldness of type.

TABLE II. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: CHILDREN

	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+	Total	Mean I.Q.
I. Higher Professional						1		1	1		3	120.8
II. Lower Professional				1	2	6	12	8	2		31	114.7
III. Clerical			3	8	21	31	35	18	6		122	107.8
IV. Skilled		1	12	33	53	70	59	22	7	1	258	104.6
V. Semiskilled	1	6	23	55	99	85	38	13	5		325	98.9
VI. Unskilled	1	15	32	62	75	54	16	6			261	92.6
Total	2	22	70	159	250	247	160	68	21	1	1000	100.0

terest among scientists. The characterization of Burt's investigation by the geneticists Cavalli-Sforza and Bodmer (*20*) is typical: "One of the most comprehensive and widely quoted studies on IQ differences between social classes and the reasons for their apparent stability is that published by Burt in 1961. His data come from school children in a typical London borough, and from their parents." Eysenck (*21*) reprinted the entire article in an anthology on the measurement of intelligence and reproduced tables I and II in two books (*22, 23*). Gottesman (*24*) also presented these tables in their entirety, as did Herrnstein in his well-known *I.Q. in the Meritocracy* (*25*). Willerman gave Burt's tables I and II in his introductory text *Individual and Group Differences* (*26*). Burt's occupational-class means and standard deviations for the fathers and children have also been widely reprinted (*20, 27–29*). Dobzhansky described the study and presented Burt's means in his *Genetic Diversity and Human Equality* (*30*).

Since the study was evidently the most comprehensive ever performed on the relation of intelligence to social class and social mobility, it is not at all surprising that it was so extensively quoted. According to Burt, "The surveys and the subsequent inquiries were carried out at intervals over a period of nearly fifty years, namely, from 1913 onwards" (*15*, p. 4). Moreover, "for the children the bulk of the data was obtained from the surveys carried out from time to time in a London borough selected as typical of the whole county" (*15*, p. 9). The adults were taken to be the fathers of the children. Kamin argues that Burt was evasive on this point and on others as well, such as sample size (*6*, pp. 153–155; *31*). The consensus, however, has been that 40,000 father-child pairs were tested (*22*, p. 62; *24*, p. 36; *26*, p. 11; *30*, p. 19; *32*, *33*). Herrnstein originally gave the sample as 1000 father-child pairs (*25*, p. 203) but later revised his estimate to 40,000, stating, "It is true that Burt's sample was 40,000, not 1,000 as I said. Burt did transform his data to a base of only 1,000 cases, which is how I made the mistake" (*33*).

Burt's occupational classification was based "not on prestige or income, but rather on the degree of ability required for the work" (*15*, p. 10). There is no mention of the intrajudge or interjudge reliability of Burt's classification procedure. According to Jensen (*29*, p. 16), Burt used the "six occupational categories of the U.S. Census."

The intelligence tests used on the children were never specified. With respect to assessment of the fathers, Burt says only: "The data for the adults was obtained from the parents of the children themselves. Usually our more immediate purpose was to secure practical estimates of both the average level and the range of intelligence required in the commoner types of occupation. In addition, however, when working with backward children we often wanted to see how far backwardness was a family characteristic. And at all levels an incidental aim was to secure material for studying the problem of mental inheritance. For obvious reasons the assessments of adult intelligence were less thorough and less reliable" (*15*, p. 9). It is clearly impossible to determine from this statement whether an intelligence test was administered (*6*). Nevertheless, let us accept Burt's description and look for cruder results for the adults. He also refers to the overall inquiry as a "pilot inquiry" and states that "the data are too crude and limited for a detailed examination by a full analysis of variance" (*15*, p. 9). Let us accept his characterization of the study and expect to find crude and limited results.

Table 3. Regression coefficient for each occupational class, computed from the class means given in Burt's tables I and II (*15*, p. 11). The regression coefficients were all rounded to two decimal places.

Class	$\bar{X}_c/(\bar{X}_f + 100)$
I	0.50
II	0.50
III	0.50
IV	0.50
V	0.50
VI	0.50

Burt's Tables I and II

The first set of data I evaluated were Burt's tables I and II (*15*, p. 11). Tables I and 2 here duplicate exactly Burt's tables I and II and his captions. Eysenck's description of those tables represents the consensus: "The data in these two tables were collected by Sir Cyril Burt on some 40,000 adults and children and have been reduced to a base of 1,000. (This means that the total of 3 for the higher professional category actually refers to 120 fathers)" (*22*, p. 62; *34*). These distributions are tabulated in section III of Burt's paper, which is entitled "Frequency distributions for adults and children" (*15*, p. 9). The motivation for the presentation of the actual distributions was that "in studying the distribution of intelligence among the different occupational classes it is in my view desirable to examine, not only (as is usually done) the class-means, but the entire frequency distributions. Accordingly in Tables I and II, I give frequencies both for adults and for children" (*15*, p. 9). In the abstract they are described as "tables compiled to show the actual distribution of intelligence among adults and children belonging to the various occupational categories" (*15*, p. 3).

Thus, Burt's tables I and II are intended to be taken as actual frequency distributions of intelligence for children and their fathers in six occupational classes, with the class arithmetic means. Burt also stated that he "rescaled our assessments of intelligence so that the mean of the whole group is 100 and the standard deviation 15" (*15*, p. 10).

In an article on class differences in intelligence (*35*), Conway gives a method for predicting the mean IQ of the children in a given class from the mean IQ of the fathers in the class. Referring to genetic regression, she says: "Thus, allowing for regression, and assuming that the I.Q. of the parents in the professional class averaged about 130, we should expect the I.Q. of their children to average about 115. Similarly, if the intelligence of the 'unskilled workers' averaged about 90, then we should expect that of their children to average about 95" (*35*, pp. 6–7). The formula is evidently $\mu_{ic} = \frac{1}{2}(\mu_{if} + 100)$, when μ_{ic} and μ_{if} are the population means of the children and fathers respectively for the i^{th} occupational class. More generally, $\mu_{ic} = \alpha(\mu_{if} + 100)$. The regression coefficient α can be estimated from $\bar{X}_{ic}/(\bar{X}_{if} + 100)$, where $\bar{X}$ denotes a sample mean.

Table 3 gives the estimates, rounded to two decimal places, of the regression coefficients that I computed from Burt's class means using the equation derived from Conway's discussion (*35*). We find that the regression is 0.50 for every class, and this agrees perfectly with Conway's prediction from the genetic theory. Since the study was described as a "pilot inquiry," the data as "crude," and the adult data as even "less thorough and less reliable," Clarke and Clarke's suspicions (*13*, p. 168) are not totally unfounded.

The Marginal Frequency Distributions

In Fig. 1 Burt's frequency distributions are pooled over occupational class. For comparison purposes I have included the theoretical normal distribution with mean 100 and standard deviation 15. It is apparent that the frequency distributions of the children and the fathers essentially coincide. The maximum absolute difference between the two distributions is found in the interval 90–100 and is 0.3 percent. Moreover, both distributions are in extraordinarily good agreement with the theoretical normal distribution. Table 4 presents Burt's column marginals for the fathers and children rounded to whole percentages. The theoretical normal distribution with mean 100 and standard deviation 15 is shown similarly. Given that Burt described his data generally as "crude and limited" (*15*, p. 9) and his adult data in particular as "less thorough and less reliable" (*15*, p. 9), the coincidence of the child and adult distributions and their perfect agreement with a normal distribution to two places are quite suspicious. Clearly, further scrutiny of these data is indicated.

The fit of the frequency distributions to a normal distribution with mean 100 and standard deviation 15 was tested by a Pearson χ^2-test of goodness-of-fit. Burt's ten classes were reduced to eight by combining the two extreme classes, 50–60 and 140+, with their respective adjacent classes. This makes excellent sense, inasmuch as the entries in the two extreme categories are of the order of the level of rounding. The X^2 for the adults was 4.889 and for the children 7.988 (*36*). Both X^2's were statistically insignificant, and the X^2 for the fathers was actually less than the theoretical expectation (*37*).

The fit of Burt's data to the theoretical normal distribution appears extraordinarily good. Moreover, it conflicts with Burt's previous findings on the fit of IQ data to the normal curve. In 1957 Burt stated (*38*, p. 167): "From 1917 onwards the normality of the frequency distributions obtained from my L.C.C. surveys was regularly checked by the χ^2 method (cf. Burt, 1917, p. 34); and in every case the distribution was found to diverge significantly from the normal." In 1963 Burt published a paper specifically concerned with the distribution of intelligence. He reported, "Frequency distributions obtained on applying intelligence tests to large samples of the school population are analysed, and compared with those given by the formulae for the commoner types of frequency curve. It is noted that the distributions actually observed are more asymmetrical and have longer tails than that described by the normal curve" (*39*, p. 175). His prescreened distribution was "plainly skewed, with a prolonged lower tail" (*39*, p. 178). I tested Burt's after-screening distribution (excess lower tail reduced) against the normal distribution using the eight intervals employed with Burt's 1961 data. The X^2 was 48.89 and highly significant ($p < 10^{-6}$) (*40*). The frequency distributions of Burt's 1961 study were not mentioned at all in his 1963 paper even though they conflict with his 1963 conclusions on the normality of IQ (*41*).

Although some may perhaps believe that the normal curve is the rule in nature, and that the findings Burt reported in his 1957 and 1963 papers are somehow exceptional, such is not the case. As was pointed out by Yule and Kendall (*42*, p. 186), "The normal curve was, in fact, to the early statisticians what the circle was to the Ptolemaic astronomers." In a discussion on the distribution of IQ, McNemar (*43*, p. 15) pointed out that "the ease with which the shape of a distribution can be altered by a change in test difficulty should also have served as a warning to those who were out to demonstrate the normal law for psychological traits." Wechsler (*44*, p. 31) remarked that "from the work that has been done the evidence is clear that, not only was Quetelet's generalization that the distribution of all or even most human traits conform to the normal law of error, premature, but incorrect." This appears to be particularly the case for adult IQ distributions. For example, Fig. 2 presents the distributions on the Army Alpha and Beta intelligence tests of the white draft of World War I (*45, 46*). The distributions are skewed. The distribution of intelligence in World War II has also been determined with the Army General Classification Test (*47*). It was found that "the distribution of test-intelligence is not normal" (*47*, p. 401). Wechsler (*48*) has presented the distribution of intelligence on the Wechsler Adult Intelligence Scale; Fig. 3 gives his distribution for a national sample of 2052 U.S. adults. The X^2 is highly significant ($p < .001$) under the normal-curve hypothesis.

Table 4. Burt's column marginals (Tables 1 and 2) as percentages rounded to whole numbers. For purposes of comparison, the theoretical normal distribution with mean 100 and standard deviation 15 [N(100, 15)] is also given.

	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130–140	140+
Adults	0	2	7	16	25	25	16	7	2	0
Children	0	2	7	16	25	25	16	7	2	0
N(100, 15)	0	2	7	16	25	25	16	7	2	0

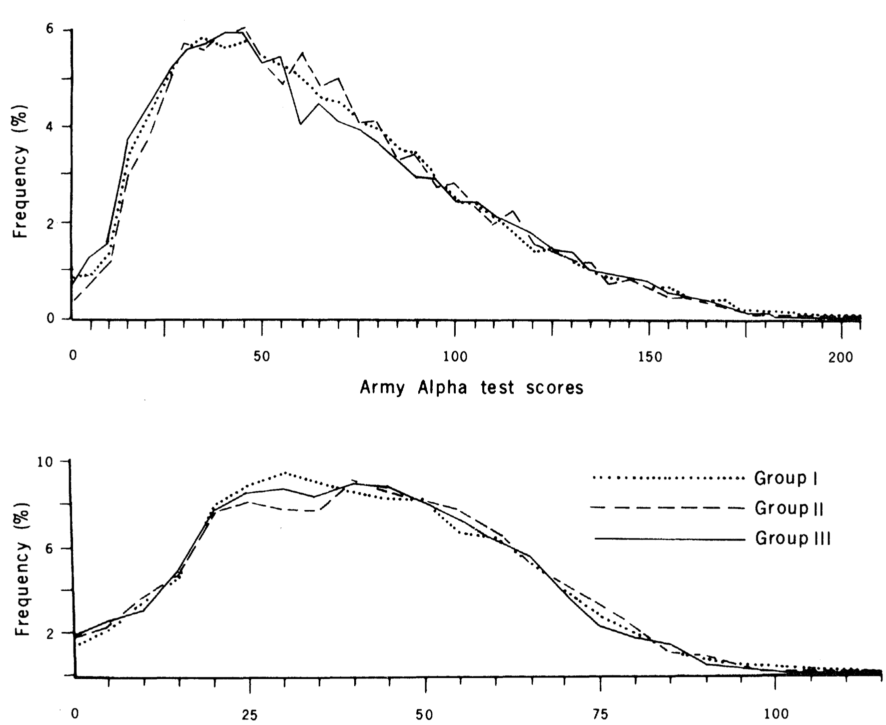

Fig. 2. Percentage frequency distributions of scores on the Army Alpha intelligence test of white draftees in groups I, II, and III who took Alpha only (upper panel) and of scores on the Army Beta of white draftees in these groups who took Beta only or Beta and Alpha (lower panel) (*45*, p. 670).

In summary, the almost perfect fit of Burt's adult and child distributions to the normal curve suggests that his "actual" (*15*, p. 3) distributions are not actual distributions.

To investigate this issue more rigorously, I fitted the normal curve to a variety of frequency distributions of intelligence, height, and weight. The fit of these distributions was compared to the fit of Burt's two distributions. To compare meaningfully distributions of different sample size, an index of goodness-of-fit was chosen which is asymptotically independent of sample size (N). I selected X^2/N. It is not hard to prove that X^2/N is asymptotically constant almost surely (a.s.). Moreover, it is easy to show that for k classes

$$\frac{X^2}{N} = \sum_{i=1}^{k} \frac{(P_{io} - P_{it})^2}{P_{it}},$$

where P_{io} is the observed proportion and P_{it} the theoretical normal proportion for the i^{th} class (*49*). Each frequency distribution was fitted to the normal curve by a computer program, which repartitioned the distribution into eight intervals as close as possible to the eight intervals used to fit Burt's 1961 data (*50*).

The results of this investigation (*51–57*) are shown in Fig. 4. The ordinate of the graph is $\log_{10} X^2/N$, and the abscissa is the common logarithm of the sample size. First, notice how close together the X^2/N's for Burt's child and adult distributions are and how distant they are from the X^2/N's of the other distributions of intelligence. But the most striking observation is that Burt's adult and child distributions gave the best and second-best fit respectively of all the distributions I tested. In fact, it may very well be that Burt's frequency distributions are the most normally distributed in the history of anthropometric measurement for intervals approximately of the form $\mu \pm (k/3)\sigma$, $k = 0, 2, 4, 6$. I invite the reader to locate a single unfabricated frequency distribution that gives a smaller X^2/N than the one obtained for Burt's adult distribution (*58*). It is interesting that the $\log_{10} X^2/N$ of Burt's 1963 study (*39*) was -1.97 for children (after screening) who were given an individually administered Stanford-Binet test as contrasted with -3.70 and -3.91 for the children and adults respectively in the "crude" 1961 data (*59*).

With the distribution of $\log_{10} X^2/N$'s from IQ used as an estimate of the null distribution, it is possible to compute an approximate probability of obtaining a $\log_{10} X^2/N$ as small as Burt's $\log_{10} X^2/N$ for the children and fathers (*60*). For

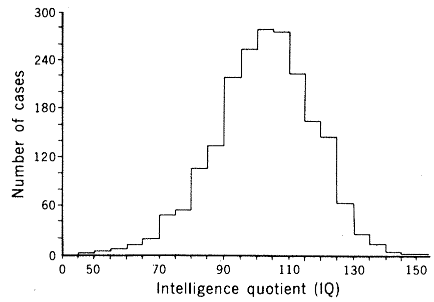

Fig. 3. Distribution of IQ's on the Wechsler Adult Intelligence Scale of a national sample of 2052 U.S. adults aged 16 and over (*48*).

Burt's fathers, $Z = 5.41$ ($p < 10^{-7}$) and for Burt's children $Z = 4.92$ ($p < 10^{-6}$). Hence, we may conclude that beyond a reasonable doubt Burt's two X^2/N's are not members of the distribution of X^2/N's for IQ.

Using the distribution of $\log_{10} X^2/N$'s for height as the null distribution, we get $Z = 3.30$ ($p < .0005$) for Burt's fathers and $Z = 2.75$ ($p < .003$) for Burt's children. Using the distribution of $\log_{10} X^2/N$'s for weight as the null distribution, we get $Z = 22.28$ ($p < 10^{-20}$) for Burt's children and $Z = 20.38$ ($p < 10^{-20}$) for Burt's fathers.

Whereas X^2/N provides an index of fit asymptotically independent of sample size, it is quite informative to compare Burt's 1961 X^2's of 4.89 and 7.99 with the X^2's of the two samples of IQ closest to 40,000, these being the 35,809 Scottish boys and 34,996 Scottish girls tested in the 1947 survey of the intelligence of almost all 11-year-old pupils in Scotland (*61*). The X^2 for the boys was 2085 and for the girls 1608. They are both highly significantly greater than Burt's X^2's (*62*). Combining the distributions of Scottish boys and girls into a single distribution and removing the 532 pupils with certain mental and physical defects (*63*), I obtained a X^2 of 3267 for the sample of 70,273. The Scottish tests were verbal. Perhaps a nonverbal test of intelligence would give a smaller X^2. A sample of 26,012 white draftees on the Army Beta, the Army test for near-illiterates in English, gave a X^2 of 1539 (*45, 64*). This X^2 was significantly greater than either of Burt's 1961 X^2's ($p < 10^{-4}$).

It would also be of interest to compare Burt's 1961 data with the classic data of Adolphe Quetelet. Quetelet was a famous 19th-century mathematician and astronomer who fathered the theory that human traits follow the normal curve (*65*). In support of his theory he presented a now classic distribution of heights of 25,878 American Civil War recruits (*65*). Truman Kelley characterized the agreement of these data with the normal curve as "striking" and asked, "Could any student . . . fail to be struck with wonder at the evidence of a permeating design?" (*66*). The X^2 for Quetelet's height data was 143.78 and significantly greater than that for Burt's children and fathers ($p < .01$).

Summing up, we may now say that, beyond a reasonable doubt, the frequen-

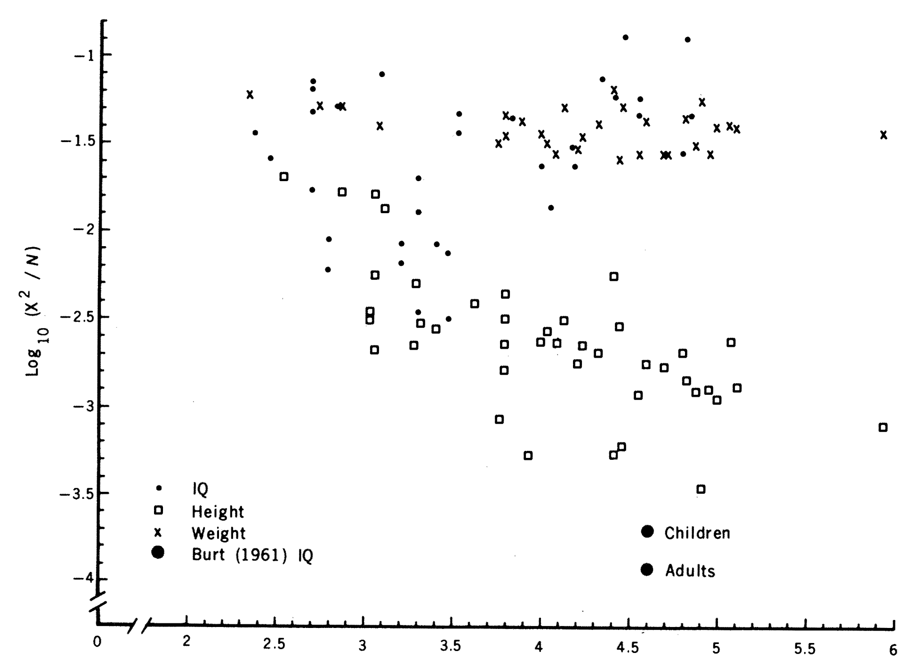

Fig. 4. $\text{Log}_{10} X^2/N$ plotted against $\log_{10} N$. The X^2/N's were computed for various frequency distributions of intelligence (*51*), height (*54*), and weight (*57*). The $\log_{10} X^2/N$'s from Burt's 1961 distributions of IQ for his children and adults (*15*) are also plotted to facilitate comparison.

cy distributions of Burt's tables I and II were carefully constructed so as to give column marginals in agreement with the normal curve. The readers were not informed anywhere in his 22-page article that that had been done.

Burt's Tables III and IV

The title of Burt's paper was "Intelligence and social mobility," and he used his tables III and IV to compute mobility across occupational classes. These computations of mobility were motivated to a large degree by some criticisms of Halsey's (*15*, p. 17, footnote 1; *67*). Halsey proposed that "the distribution of innate intelligence among different social classes is approximately random" (*67*, p. 1) and that "with more precise figures for social mobility, their [Conway and Burt's] arguments would strengthen rather than weaken the hypothesis here maintained [the environmental hypothesis]" (*67*, p. 1). Burt's 1961 computations disposed of Halsey's criticism. Burt concluded, "The foregoing data and the analysis I have here attempted will, I hope, dispose of one of the strongest objections urged by Dr. Halsey (21) [Burt's reference; here it is (*67*)] against the arguments brought forward by Miss Conway and myself" (*15*, p. 17).

To compute results on mobility, Burt reclassified the "actual data" (*15*, p. 12) according to a new set of IQ intervals or categories using a new set of category border lines. Thus, in introducing his table III he said: "If we now reclassify the actual data for adults according to these new borderlines, we obtain the distribution set out in Table III" (*15*, p. 12). His table IV gives the results of the reclassification of the children's data. Tables 5 and 6 here duplicate Burt's tables III and IV respectively (*68*).

The new category border lines were obtained by converting each cumulative row proportion,

$$P_k = \sum_{i=1}^{k} P_{i\cdot}$$

to the corresponding inverse value of the normal distribution function, N(0, 1), where $P_{i\cdot}$ is the marginal proportion of the *i*th row or, equivalently, of the *i*th occupational class. Table 7 gives useful information on the conversion of P_k to category border lines. Note that Z^* is the inverse of the normal distribution function, N(0, 1), and Z^{**} is the Z-score that gives Burt's border line (*69*). Since Burt evidently rounded, Z^* is identical to Z^{**} only on the first two figures.

In Burt's tables III and IV we see that the column marginals are almost identical for the children and the adults. Moreover, with category 141+ pooled with its adjacent category (*70*), the column marginals are identical for the children and the adults, a match that seems extraordinary for "crude and limited" data (*15*, p. 9). Indeed, $X^2 = 0$ for both the children and the adults under the null hypothesis that the row marginals equal the column marginals after categories I and II are pooled. The probability of obtaining a X^2 less than 10^{-6} by chance is smaller than two in 1 trillion for 4 degrees of freedom. In discussing small X^2's, Yule and Kendall point out (*42*, p. 423) that "very close correspondence is *too good to be true*" (their italics). Thus, beyond a reasonable doubt, Burt's assertion that "if we now reclassify the actual data for adults according to these new borderlines, we obtain the distribution set out in Table III" (*15*, p. 12) is fraudulent. His X^2's of zero are too good to be true. It is also bizarre that Burt's data are no longer normally distributed with mean 100 and standard deviation 15. Under this hypothesis (*71*) the Pearson X^2 is 84.99 for both the children and the adults and is highly significant ($p < 10^{-9}$, d.f. = 4). But it is clear from Table 7 that the discrepancy is a result of the fact that Burt's new border lines are from rounded Z-scores. Indeed, the assumption of a normal distribution, N(100, 15), for the column marginals does give X^2's of zero for the children and the adults (*71*) if we use category border lines X^* from the unrounded Z-scores (see Table 7). But Burt used border lines X^{**} from rounded Z-scores to reclassify "the actual data." In brief, beyond any reasonable doubt, Burt did not reclassify actual data, as he asserted, but constructed the distributions. I invite the reader to devise an alternative rigorous explanation for the X^2's of zero. An alternative explanation must account for X^2's of zero for both the adults and the children under the assumption that "the assessments of adult intelligence were less thorough and less reliable" (*15*, p. 9).

Burt's column marginals were elegantly used in his computations of mobility through his tables V and VI (*15*, p. 13). Thus those tables, which were supposed to "dispose" of Halsey's criticism, were completely determined from tables III and IV.

Table 5. Burt's table III (*15*, p. 13), copied exactly except for boldness of type.

TABLE III. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: ADULTS

				Rescaled			
	VI	V	IV	III	II	I	Total
	50–91	91–103	103–115	115–127	127–141	141+	
I					2	1	3
II			1	15	14	1	31
III	1	15	38	56	12		122
IV	16	86	114	38	4		258
V	53	178	84	10			325
VI	191	46	21	3			261
Total	261	325	258	122	32	2	1000

Table 6. Burt's table IV (*15*, p. 13), copied exactly except for boldness of type.

TABLE IV. DISTRIBUTION OF INTELLIGENCE ACCORDING TO OCCUPATIONAL CLASS: CHILDREN

				Rescaled			
	VI	V	IV	III	II	I	Total
	50–91	91–103	103–115	115–127	127–141	141+	
I			1	1	1		3
II	1	4	11	9	6		31
III	11	28	51	20	12		122
IV	46	66	75	62	8	1	258
V	91	122	84	23	5		325
VI	112	105	36	7	1		261
Total	261	325	258	122	33	1	1000

Burt's Row Totals

The row totals of Burt's tables I to IV and the column totals of his tables III and IV would appear on the basis of Burt's descriptions and discussions to be simply totals per mille. Moreover, such is the conclusion of experts in the field

(*20, 22, 24–26, 28, 30, 33*). Thus, in discussing tables I and II Burt said (*15*, pp. 14–15): "Consider, for example, the lowest occupational class of all. Among the adults only 20 persons out of 261 have an intelligence above the general average; among the children as many as 76, nearly four times as many—a discrepancy of 56. . . . Here again the increased numbers would be almost inexplicable on the environmental theory, but a natural consequence of the Mendelian theory of polygenic inheritance." A little later he says, "Our method of reducing the figures observed to numbers per 1000 should sufficiently allow for the change in the absolute size of the population" (*15*, p. 16). In short, according to Burt, the figures observed were reduced to numbers per 1000. It follows that the row totals were the figures observed reduced to numbers per 1000.

There exist figures associated with "vocational categories," published by Spielman and Burt in 1926 (*72*), that bear an extraordinary resemblance to the row totals of Burt's tables I to IV and the column totals of his tables III and IV. Table 8 gives the percentage of male adults estimated in each vocational category by Spielman and Burt in their table III (*72*, p. 13); their percentages for categories 1 and 8 have been rounded to whole numbers. In the next column to the right, I have put Burt's 1961 row totals as percentages similarly rounded. Burt's 1961 totals agree perfectly with Spielman and Burt's totals for the first five levels. Spielman and Burt (*72*) placed "casual labour" in level 7, whereas Burt (*15*) placed "casual labourers" in level 6 and had no lower levels. Hence, let us pool Spielman and Burt's levels 6 and 7 (see Table 8) into a sixth level. There is now perfect agreement for all six levels. Such a coincidence is highly unlikely by chance. The probability of getting an exact match on each of the six percentages is 1 in 97 million (*73*). Thus, beyond any reasonable doubt, Burt's (*15*) row and column totals of tables III and IV and his row totals of tables I and II came from Spielman and Burt's percentages published in 1926, with the addition of a decimal place. Since Burt's surveys were "carried out at intervals over a period of nearly fifty years, namely, from 1913 onwards" (*15*, p. 4) and his 1961 data were relatively "crude and limited" (*15*, p. 9), the coincidence is bizarre indeed.

Perhaps Spielman and Burt's percentages were computed from a representative sample of fathers from the London borough described by Burt as "typical of the whole county" (*15*, p. 9; *74*). This conjecture can be rejected. First, they stated that "the classification and the percentages have been taken, with some modification and revision, from the table printed by Burt in a paper on 'Some Principles of Vocational Guidance' (*Brit. Journ. Psychol.*, 1924, III, p. 349)" (*72*, p. 15, footnote 1). In the 1924 paper Burt reported percentages for each vocational category (*75*), giving their origin thus: "For London a rough calculation (based mainly upon Charles Booth's survey, corrected by later census figures) gives the following proportions" (*75*, p. 349; *76*). Concerning their final figures, Spielman and Burt state (*72*, p. 15, footnote 1): "Once more the figures finally arrived at are to be taken as nothing more than the roughest approximation. The propor-

Table 7. Numerical transformations from proportions in each occupational class ($P_{i\cdot}$) to new category border lines (X^{**}) for Burt's tables III and IV (*15*, p. 13). Note that $X^* = -15Z^* + 100$ and Z^{**} is the solution to $Z^{**} = (100 - X^{**})/15$. Burt began with Z^* and apparently rounded to Z^{**} to obtain X^{**}. F is the normal distribution function, N(100, 15).

Class	k	$P_{i\cdot}$	$\sum_{i=1}^{k} P_{i\cdot}$	Z^*	Z^{**}	X^{**}	X^*	$F(-X^{**})$	$F(-X^*)$
I	1	.003	.003	−2.7478	$-2.7\frac{1}{30}$	141	141.2170	.0031	.0030
II	2	.031	.034	−1.8250	−1.8	127	127.3750	.0359	.0340
III	3	.122	.156	−1.0110	−1.0	115	115.1650	.1587	.1560
IV	4	.258	.414	−0.2173	−0.2	103	103.2595	.4207	.4140
V	5	.325	.739	+0.6403	+0.6	91	90.3955	.7257	.7390
VI	6	.261	1.000	+∞	+∞			1.0000	1.0000

Table 8. Percentages of male adults in the vocational categories given in Spielman and Burt (*72*, table III, p. 13). The percentages in levels 1 and 8 were rounded to whole numbers (the original figures are in parentheses). For comparison purposes, the percentage of adults in the six occupational classes in Burt's 1961 paper (*15*) are similarly rounded.

Category level	Definitions: Spielman and Burt (*72*)* (vocational category)	Definitions: Burt (*15*)† (occupational class)	Percentage: Spielman and Burt (*72*)	Percentage: Burt (*15*)
1	Highest professional and administrative work	"highest type of professional and administrative work"	(0.1) 0	0
2	Lower professional and technical work	"lower professional or technical work"	3	3
3	Clerical and highly skilled work	"intermediate types of clerical, commercial, or technical work"	12	12
4	Skilled work. Minor commercial positions	"skilled workers. . . . commercial or industrial work of an equivalent level"	26	26
5	Semi-skilled work. Poorest commercial positions	"semi-skilled workers and those holding the poorest type of commercial position"	33	33
6	Unskilled labour and coarse manual work	"unskilled labourers, casual labourers, and those employed on coarse manual work"	19 } 26	26
7	Casual labour		7 } 26	—
8	Institutional cases (imbeciles and idiots)		(0.2) 0	—

*These definitions are precisely as given by Spielman and Burt (*72*, p. 13) in their column 4 labeled "Vocational category." †Taken from the text of Burt's 1961 paper (*15*, p. 10).

tions have been computed primarily from the figures given in the Census returns for London [(*77*)]. . . . Approximate as they are, we feel that these figures are still worth recording. A rough numerical guide is better than no guide at all.'' Thus, Spielman and Burt's percentages were a rough estimate of the ''percentage of persons following trades or professions belonging to each group or class'' (*72*, p. 15) up to 1921, the last English census before 1926. Hence Burt's ''row totals'' were based upon a census taken nearly 40 years before completion of his 1961 study (*15*, p. 4; *78*). Moreover, Spielman and Burt's percentages were not of male adults with at least one child from a specific London borough. Since the probability of an exact match by chance to just the units digit of the six percentages of Spielman and Burt is 1 in 100,000, it is beyond any reasonable doubt that Burt's 1961 row totals were not the observed row totals per mille. Beyond any reasonable doubt, they were taken by Burt from the Spielman and Burt study.

Conclusions

These findings show, beyond any reasonable doubt, that Burt fixed the row and column totals of the tables in his highly acclaimed 1961 study. Since the totals are completely determined by the cell entries, Burt determined the cell entries. Thus Burt's so-called ''actual'' frequency distributions were systematic constructions. Burt discussed the constructions as if they were observed frequency distributions per mille; hence it is not surprising that authorities in the field were misled (*20, 22–30, 33*). Burt never said that he fixed *both* the column and the row totals of his tables, and he gave no clues to the algorithm by which the cell entries were computed subject to constraints on both row and column totals. Moreover, Burt used his systematic constructions to test deductions from his genetic theory of social class and to answer criticisms of previous work. Since constructions are not data, it is extraordinary for Burt to have discussed his tables. as if they were actual data. Table 9 illustrates ways in which he did so.

It is also extraordinary that Burt ''was a deadly critic of other people's work when this departed in any way from the highest standards of accuracy and logical consistency'' and that ''he could tear to ribbons anything shoddy or inconsistent'' (*79*, p. iv).

The final judgment on Burt's honesty as a scientist will rest with future historians of behavioral science. It is hoped that the foregoing analyses will contribute to a fair and reasoned judgment in this matter.

Table 9. Quotations from ''Intelligence and social mobility'' (*15*) showing that Burt discussed his constructions as if they were data and used the constructions to confirm his theory and answer criticisms.

''These deductions are fully confirmed by tables compiled to show the actual distribution of intelligence among adults and children belonging to the various occupational categories'' [p. 3 (abstract)].

''I propose to offer more detailed evidence to support the interpretations I put forward'' (p. 3).

''The data which I shall analyse are drawn from two overlapping inquiries'' (p. 3).

''. . . much of the data is due to the willing cooperation of numerous collaborators'' (p. 4).

''*Sources of Data* [italics in original]. In studying the distribution of intelligence among the different occupational classes it is in my view desirable to examine, not only (as is usually done) the class-means, but the entire frequency distributions. Accordingly in Tables I and II, I give frequencies both for adults and for children. For the children the bulk of the data was obtained from the surveys carried out from time to time in a London borough. . . . The data for the adults was obtained from the parents of the children themselves'' (p. 9).

''If we now reclassify the actual data for adults according to these new borderlines, we obtain the distribution set out in Table III'' (p. 12).

''When we turn to the data for children (Table II)'' (p. 12).

''Among the adults only 20 persons out of 261 have an intelligence above the general average; among the children as many as 76, nearly four times as many—a discrepancy of 56'' (p. 14).

''On the Mendelian hypothesis, however, such apparent anomalies are exactly what we should anticipate'' (p. 15).

''In the first three occupational classes, for example, we see that among the adults only 9 out of 156 had an intelligence below the general mean, among the children as many as 39. Here again the increased numbers would be almost inexplicable on the environmental theory, but a natural consequence of the Mendelian theory of polygenic inheritance'' (p. 15).

''Our present data supply us with two such samples. These are comparable since the adults are the parents of the children'' (p. 16).

''Assuming then that the data in our two samples are reasonably comparable'' (p. 17).

''The foregoing data and the analysis I have here attempted will, I hope, dispose of one of the strongest objections urged by Dr. Halsey'' (p. 17, footnote 1).

''Owing to the imperfect correlation between the intelligence of parents and the intelligence of their children the discrepancies between the children's intelligence and the occupational category of the parents are still greater. This follows from the multifactorial theory of inheritance, and is amply confirmed by the data here examined'' [p. 23 (Summary and Conclusions)].

References and Notes

1. A. R. Jensen, *Psychometrika* **37**, 115 (1972).
2. O. Gillie, *Sunday Times* (London), 24 October 1976, p. 1.
3. ——, *Phi Delta Kappan* **58**, 469 (1977).
4. B. Rensberger, *N.Y. Times*, 28 November 1976, p. 26.
5. N. Wade, *Science* **194**, 916 (1976).
6. L. Kamin, *The Science and Politics of I.Q.* (Erlbaum, Potomac, Md., 1974).
7. C. Burt, *Br. J. Educ. Psychol.* **25**, 158 (1955).
8. ——, *Am. Psychol.* **13**, 1 (1958).
9. ——, *Br. J. Psychol.* **57**, 137 (1966).
10. B. Rimland and H. Munsinger, *Science* **195**, 248 (1977).

→ D. W. Fulker, *Am. J. Psychol.* **88**, 505 (1975).

12. J. Cohen, *Times* (London), 10 November 1976, p. 17.
13. A. M. Clarke and A. D. B. Clarke, in *Mental Deficiency: The Changing Outlook*, A. M. Clarke and A. D. B. Clarke, Eds. (Methuen, London, 1974).
14. M. McAskie and A. M. Clarke, *Br. J. Psychol.* **67**, 243 (1976). This is a comprehensive review of parent-offspring resemblances in intelligence. The findings are quite variable. The degree of parent-child resemblance appears to depend upon a number of factors, such as degree of similarity between the IQ tests administered to the children and to their parents, the reliability of the test scores, and the ages of and age differences between the parents and children at the time of test administration.
15. C. Burt, *Br. J. Stat. Psychol.* **14**, 3 (1961).
16. ——, *Br. J. Educ. Psychol.* **13**, 83 (1943).
17. A. R. Jensen, *Behav. Genet.* **4**, 1 (1974).
18. ——, *Phi Delta Kappan* **58**, 471 (1977), reprinted from the *Times* (London), 9 December 1976, p. 11. This is Jensen's latest view as of 4 April 1977 (personal communication).
19. H. J. Eysenck, *Am. Psychol.* **32**, 674 (1977).
20. L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations* (Freeman, San Francisco, 1971), p. 795.
21. H. J. Eysenck, *The Measurement of Intelligence* (Williams & Wilkins, Baltimore, 1973), pp. 362–383.
22. ——, *The I.Q. Argument: Race, Intelligence, and Education* (Library Press, New York, 1971), p. 62.
23. ——, *The Inequality of Man* (Temple Smith, London, 1973), p. 139.
24. I. I. Gottesman, in *Social Class, Race, and Psychological Development*, M. Deutsch, I. Katz, A. R. Jensen, Eds. (Holt, Rinehart & Winston, New York, 1968), pp. 36–37.
25. R. J. Herrnstein, *I.Q. in the Meritocracy* (Little, Brown, Boston, 1973), p. 204.
26. L. Willerman, *Individual and Group Differences* (Harper & Row, New York, 1975), pp. 12–13.
27. W. F. Bodmer and L. L. Cavalli-Sforza, *Genetics, Evolution, and Man* (Freeman, San Francisco, 1976), p. 682.
28. L. Ehrman and A. Parsons, *Genetics of Behavior* (Sinauer, Sunderland, Mass., 1976), p. 293.

→ A. R. Jensen, *Am. Educ. Res. J.* **5**, 1 (1968).

30. Th. Dobzhansky, *Genetic Diversity and Human Equality* (Basic Books, New York, 1973), p. 20.
31. In the paper (*15*) Burt usually referred to the subjects as adults and children. Nonetheless, the most reasonable interpretation is that they were fathers and sons. He used the terms ''fathers and sons'' in the abstract (*15*, p. 3) and intermittently used the words ''boys'' (pp. 9 and 14) and ''fathers and sons'' (p. 14). In discussing tables I and II he said, ''When we look at the distribution of the children's intelligence . . . and compare it with that of their fathers' (Table I) . . .'' (*15*, p. 14).

32. J. Hirsch, *Comtemp. Psychol.* **20**, 436 (1975).
33. R. J. Herrnstein, *ibid.*, p. 436.
34. As stated by Burt, "Our present data supply us with two such samples. . . . Our method of reducing the figures observed to numbers per 1000 . . ." (*15*, p. 16).
35. J. Conway, *Br. J. Stat. Psychol.* **12**, 5 (1959).
36. The normal distribution was computed from Hastings's most accurate rational approximation [C. Hastings, Jr., *Approximations for Digital Computers* (Princeton Univ. Press, Princeton, N.J., 1955), p. 187]. The absolute error of approximation is less than 1.5×10^{-7}.
37. The distribution of X^2 for this case is bounded between a χ^2_5 and χ^2_7 variable [M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Hafner, New York, 1961), vol. 2, p. 430]. Also, it can be shown that rounding the observed frequencies to per mille increases the mean X^2.
38. C. Burt, *Br. J. Psychol.* **48**, 161 (1957).
39. ———, *Br. J. Stat. Psychol.* **16**, 175 (1963).
40. The theoretical normal distribution was computed from Burt's reported mean of 100.07 and standard deviation of 15.09.
41. As evidence against the view that illness prevented Burt from recomputing his twin correlations, in this study he performed a laborious fit of the frequency distribution to a Pearson type IV curve. He declared, "We have found the labour lengthy rather than difficult" (*39*, p. 181). The recomputation of his twin correlations would seem to be a more important project.
42. G. U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics* (Griffin, London, 1937).
43. Q. McNemar, *The Revision of the Stanford-Binet Scale. An Analysis of the Standardization Data* (Houghton Mifflin, Boston, 1942).
44. D. Wechsler, *The Range of Human Capacities* (Hafner, New York, 1952).
45. R. M. Yerkes, Ed., *Psychological Examining in the United States Army*, Memoirs of the National Academy of Sciences (Government Printing Office, Washington, D.C., 1921).
46. My Fig. 2 comes from (*45*, figures 10 and 11, p. 670). Group I represented the white draft at large, group II weighted each state more evenly, and group III weighted each camp more evenly.
47. Staff, Personnel Research Section, *J. Educ. Psychol.* **38**, 385 (1947).
48. D. Wechsler, *The Measurement and Appraisal of Adult Intelligence* (Williams & Wilkins, Baltimore, 1958), p. 107. My Fig. 3 comes from Wechsler's figure 4, p. 107.
49. It is not difficult to show that if the normal-curve hypothesis is false, then $X^2 \rightarrow +\infty$ a.s. as $N \rightarrow \infty$, whereas X^2/N converges a.s. whether the null hypothesis is true or false (assuming that $P_{it} \neq 0$). Moreover, X^2/N can be computed directly from the observed proportions (P_{io}) if they are given. Thus, X^2/N was computed from Burt's frequency distributions as represented by his column totals divided by 1000. Hence, Burt's sample size was not used to compute X^2/N for his data. It should be emphasized that under the normal curve hypothesis, N(100, 15), rounding the f_{io} to per mille increases the mean X^2/N by a small constant. A discussion of the index, X^2/N, is given by Y. M. M. Bishop, S. E. Fienberg, P. W. Holland, in *Discrete Multivariate Analysis: Theory and Practice*, Y. M. M. Bishop *et al.*, Eds. (M.I.T. Press, Cambridge, Mass., 1975), pp. 329–332.
50. The program, called NORMFIT, was written to fit the frequency distributions to the normal curve. It computed P_{io}, P_{it}, and the index of fit

$$\sum_{i=1}^{8} \frac{(P_{io} - P_{it})^2}{P_{it}} .$$

The program was written in double precision and used Hastings's most accurate approximation to the normal distribution for the computation of the P_{it} (*36*). If the sample mean ($\bar{X}$) and the sample standard deviation (S) were reported, they were used to compute the P_{it}. Otherwise, $\bar{X}$ and S were computed from the frequency distribution. Recall that I used $\bar{X} = 100$ and $S = 15$ to compute the theoretical proportions for Burt's data.

The program repartitioned each distribution into eight categories (intervals) by selecting a subset of seven category boundaries (b_i, $1 \leq i \leq 7$) from the distribution's original set of boundaries. The boundaries were selected to be as close as possible to Burt's boundaries in standardized form. Thus, if $\bar{X}$ and S are the mean and standard deviation of the distribution in question, the seven boundaries were selected to be as close as possible to $\bar{X} - 2S$, $\bar{X} - (4/3)S$, $\bar{X} - (2/3)S$, $\bar{X}$, $\bar{X} \pm (2/3)S$, $\bar{X} + (4/3)S$, $\bar{X} + 2S$. Hence b_1 was the boundary closest to $\bar{X} - 2S$; b_2 was the boundary closest to $\bar{X} - (4/3)S$, and so on, where $b_1 < b_2 < \ldots < b_7$. In brief, the program pools adjacent categories so that the new categories are as close to Burt's as possible in standardized form. The normal curve was fitted to this repartitioned distribution. Note that there is no a priori reason why category boundaries of the form $[\bar{X} \pm (k/3)S$, $k = 0, 2, 4, 6]$ should favor Burt's frequency distributions with respect to the fit to the normal curve.
51. Burt's studies "were carried out at intervals over a period of nearly fifty years, namely, from 1913 onwards" (*15*, p. 4). Since Burt did not specify any of the tests he used, I sampled frequency distributions of intelligence of other investigators from the entire time period of Burt's surveys. I chose distributions based on moderately large to very large samples to facilitate comparison.

I fitted the following 33 frequency distributions of intelligence test scores to the normal curve: (i) Four frequency distributions of scores on the Army Group Intelligence Examination A, the predecessor of the Army Alpha Examination; the four samples were 15,270, 11,380, 21,924, and 15,073 white, enlisted, literate men from camps Devens, Dix, Lee and Taylor, respectively, tested in 1917; I used the proportionate frequency distributions given in units of 0.10 percent (convertible to frequencies per mille) presented in (*45*, table 146, p. 492); I also fitted the total frequency distribution of the four camps, which has a sample size of 63,647 (*45*, table 146, p. 492). (ii) A frequency distribution of scores on the Army Alpha; the sample was 29,519 white draftees who had taken the Alpha alone and who were from group I, the sample developed to represent the U.S. draftee population at large, tested in 1918; the distribution was taken from (*45*, table 174, p. 659). (iii) A frequency distribution of scores on the Alpha; the sample was 67,254 white draftees from groups I, II, and III who were tested in 1918 and who had taken the Alpha alone (*45*, table 183, p. 666). (iv) A distribution of mental ages from the individually administered Stanford-Binet; the sample was 1246 white draftees from groups I, II, and III tested in 1918 (*45*, table 185, p. 667). (v) A distribution of scores on the individually administered Yerkes-Bridges point scale; the sample was 689 white draftees from groups I, II, and III tested in 1918 (*45*, table 186, p. 667). (vi) A distribution of scores on the Army Beta, the group test for foreigners or near-illiterates in English; the sample was 26,012 white draftees from groups I, II, and III tested in 1918 (*45*, table 189, p. 669). (vii) A proportionate frequency distribution (in units of .01 percent) of scores on the Beta; the sample was 10,027 men from group I tested in 1918 (*45*, table 192, p. 671). (viii) Two distributions of scores on the original Stanford-Binet intelligence test and two distributions of verbal scores on the Scottish Group Test of Intelligence [Scottish Council for Research in Education, *The Intelligence of Scottish Children* (Univ. of London Press, London, 1933), p. 92]; the two samples for both tests were 500 Scottish boys and 500 Scottish girls selected to be representative of all Scottish children born in 1921 and tested in 1932. (ix) Two distributions of scores on the original Stanford-Binet [D. G. Paterson, in (*52*), table 28, p. 167]; the samples were 237 children with diseased tonsils and 294 children with normal tonsils. (x) Six distributions of IQ scores on the 1937 revised Stanford-Binet [Q. McNemar, *The Revision of the Stanford-Binet Scale* (Houghton Mifflin, Boston, 1942), table 1, p. 21; table 2, p. 22]; one distribution of IQ's on form L and one on form M for 1623 children 6 to 13 years of age; one distribution on form L and one on form M for 619 adolescents 14 to 18 years of age; one distribution of IQ's on form L and one on form M for 2970 children 2½ to 18 years of age; the samples consisted of the subjects from the standardization group in the specified age ranges; the standardization group was equally divided between the sexes, white, and American born, and selected to represent "the white child population of the main geographical areas of the United States" (p. 2). (xi) A distribution of 2553 index-of-brightness scores on the Advanced Otis Scale, form A; the sample was almost all children whose homes were in the city of Bath, England, on 27 July 1934 and whose ages were between 9 and 13 years inclusive [J. A. Fraser Roberts, R. M. Norman, R. Griffiths, *Ann. Eugen.* **6**, 319 (1935), appendix table]. (xii) Two distributions of verbal test scores on the Scottish Group Test of Intelligence; the samples were 35,809 boys and 34,996 girls, who constituted almost the entire population of 11-year-old Scottish children in 1947, the year the test was administered; 532 children who were unable to attempt the test because of physical or mental defect (group YY) were assigned zero by the researchers [Scottish Council for Research in Education, *The Trend of Scottish Intelligence* (Univ. of London Press, London, 1949), table 26, p. 125]. (xiii) The distribution of verbal test scores on the Scottish Group Test of Intelligence for 70,273 Scottish 11-year-olds, the aggregate of the two samples described just above, excluding the 532 YY children (*53*, table 1, p. 200). (xiv) Three distributions of verbal test scores on the Scottish Group Test of Intelligence from 3395 boys, 3415 girls, and the aggregate sample of 6810 boys and girls; the sample was a subset of the Scottish 11-year-olds (excludes YY) described above, consisting of those in the 1947 survey born on the first three days of each month in 1936 (*53*, table 2, p. 203). (xv) Three distributions of 2052 IQ scores, one on the verbal scale, one on the performance scale, and one on the full scale of the Wechsler Adult Intelligence Scale (WAIS); the sample was a national standardization group of 1700 subjects from ages 16 to 64, and 352 subjects from Kansas City aged 60 years and over [D. Wechsler, *The Measurement and Appraisal of Adult Intelligence* (Williams & Wilkins, Baltimore, 1958), table 73, p. 253].
52. J. A. Harris, C. M. Jackson, D. G. Paterson, R. E. Scammon, Eds., *The Measurement of Man* (Univ. of Minnesota Press, Minneapolis, 1930).
53. Scottish Council for Research in Education, *Social Implications of the 1947 Scottish Mental Survey* (Univ. of London Press, London, 1953).
54. I selected a diversity of frequency distributions of height. Some were chosen because they were explicitly characterized by authors as good examples of the normal curve. The following 42 distributions were fitted to the normal curve: (i) A distribution of the heights of 25,878 American Civil War recruits [A. Quetelet, *Anthropométrie ou Mesure des Différentes Facultés de l'Homme* (Muquardt, Bruxelles, 1870), table on p. 259]. (ii) A compound distribution of the heights of 8585 adult males born in the British Isles and the four component distributions of these adult males born in England (N = 6194), Scotland (N = 1304), Wales (N = 741), and Ireland (N = 346); the original data appeared in an 1883 final report of the Anthropometric Committee to the British Association (*42*, table 6.7, p. 94). (iii) A distribution of heights of 1079 fathers and the 1079 mothers [K. Pearson and A. Lee, *Biometrika* **2**, 357 (1903), table 13, p. 408]. (iv) A distribution of heights of 2501 English male convicts [J. A. Harris, in (*52*), table 6, p. 43]. (v) Twenty-two frequency distributions of heights taken from the first 1 million U.S. Army draft recruits of World War I who were between the ages of 21 and 30 years inclusive. The 22 groups were constituted on the basis of ethnic, occupational, physiographic, and racial criteria with some overlap among some distributions (*55*, table 23, p. 109). The groups were 66,885 recruits from geographical sections of the agricultural North with native whites constituting over 73 percent of the local population (group No. 1); 97,338 recruits from sections of the agricultural North and West with similar proportions of foreign and native white local populations (group No. 2); 117,548 recruits from native white sections of the agricultural South (group No. 3); 49,503 recruits from sections of the agricultural South with populations at least 45 percent black (group No. 4); 81,718 recruits from Eastern manufacturing sections (group No. 5); 29,032 recruits from commuter-suburban areas (group No. 6); 35,730 recruits from mining sections (group No. 7); 16,165 recruits from sparsely settled sections having not more than three persons per square mile (group No. 8); 6121 recruits from desert sections (group No. 9); 6161 recruits from maritime sections (group No. 10); 17,099 recruits from mountainous sections (group No. 11); 21,254 recruits from sections primarily inhabited by mountain whites of southeastern United States (group No. 12); 10,038 recruits from sparsely settled sections with relatively large American Indian populations (group No. 13); 10,779 recruits from sparsely settled areas with relatively large Mexican populations (group No. 14); 13,522 recruits from sections with local populations characterized as native white of Scotch origin (group No. 15); 12,076 recruits from sections with at least 10 percent of the population of Russian origin (group No. 16); 51,009 recruits from sections with at least 10 percent of the population of Scandinavian origin (group No. 17); 5864 recruits from sections with at least 10 percent of Finnish origin (group No. 18); 25,862 recruits from sections with at least 10 percent of French Canadian origin (group No. 19); 28,095 recruits

from sections with at least 10 percent of German or Scandinavian origin (group No. 20); 38,962 recruits from sections with at least 20 percent of German or Austrian origin (group No. 21); 126,994 recruits from sections with at least 15 percent of German or Austrian origin, where this group contains the preceding group (group 22). (vi) A frequency distribution of the heights of 868,445 draft recruits from the first 1 million U.S. Army draft recruits of World War I (*55*, table I, appendix, p. 417). (vii) Three frequency distributions of heights of U.S. troops of World War I measured at demobilization (1919), a distribution of 4204 soldiers of English origin, a distribution of 2074 soldiers of Scotch origin, and a distribution of 6164 soldiers of Irish origin (*55*, table 25, p. 116). (viii) A distribution of heights of 91,161 males between the ages of 20 and 21, consisting of essentially all British subjects required to register for military training in 1939, and a distribution of heights of 74,489 males consisting of the subset of this sample with medical grade I (full normal health and strength) (*56*, table 17, p. 65). (ix) A distribution of heights of 1959 males and a distribution of heights of 1908 females selected to be representative of the adult population of the United States, especially those who habitually used the railroads [E. A. Hooton, *A Survey in Seating*, (Heywood-Wakefield Co., Gardner, Mass., 1945), tables on p. 39 and p. 41]. (x) Three frequency distributions of heights, one from 1162 Vassar students, one from 1162 fathers, and one from 1162 mothers of the students [M. L. Hathaway and E. D. Foard, *Heights and Weights of Adults in the United States*, Home Economics Report No. 10, Department of Agriculture (Government Printing Office, Washington, D.C., 1960), table 28, p. 40].

55. C. B. Davenport and A. G. Love, Medical Department of the United States Army, *Army Anthropology* (Government Printing Office, Washington, D.C., 1921).
56. W. J. Martin, *The Physique of Young Adult Males*, Med. Res. Coun. (G.B.) Memo. No. 20 (H.M. Stationery Office, London, 1949).
57. The following 30 distributions of weight were fitted to the normal curve: (i) A compound distribution of weights of 7749 adult males born in the British Isles, and the four component distributions of adult males born in England (N = 5552), Scotland (N = 1212), Wales (N = 738), and Ireland (N = 247) (*42*, table, p. 111). (ii) Twenty-two frequency distributions of weight taken from the first 1 million U.S. Army draft recruits of World War I, constituted on the basis of ethnic, occupational, physiographic, and racial criteria with some overlap among some distributions (*55*, table 35, p. 133); a frequency distribution of weights of 868,445 draft recruits from the first 1 million U.S. Army draft recruits of World War I (*55*, table I, appendix, p. 417). (iii) A distribution of the weights of 91,084 males between the ages of 20 and 21, consisting of essentially all British subjects required to register for military training in 1939, and a distribution of weights of 74,429 males consisting of the subset of this sample with medical grade I (*56*, table 16, p. 64).
58. The computer program, NORMFIT, is available upon request.
59. The mean of these two $\log_{10} X^2/N$'s is −3.81, and the standard deviation is 0.1485. Burt's after-screening 1963 $\log_{10} X^2/N$ is over 12 standard deviations above this mean. Note that Burt's before-screening distribution was "plainly skewed, with a prolonged lower tail" (*39*, p. 178). The screening removed some of this lower tail.
60. Assume a sample of size m is taken from a population of frequency distributions and that the random variable $\log_{10} X^2/N$ is approximately normally distributed in this population with mean μ and variance σ^2. Compute the sample mean $\bar{X}_m$ and sample variance S^2_m. Assume that $(\bar{X}_m, S^2_m)$ converges in probability to (μ, σ^2), which is weaker than the usual assumption that the random elements of a sample be mutually independent. Let X be a new observation. Under the null hypothesis that X comes from this population, the statistic $(X - \bar{X}_m)/S_m$ is approximately asymptotically normally distributed, N(0, 1). In support of the use of this approximation neither the skew ($\sqrt{\beta_1}$) nor the kurtosis (β_2) was significantly different from the normal in the distributions of $\log_{10} X^2/N$ for intelligence, height, and weight [for the test, see C. R. Rao, *Advanced Statistical Methods in Biometric Research* (Wiley, New York, 1952), pp. 218–219].
61. See (*51*, item xii) for details.
62. Let λ_0 and λ_1 be the noncentrality parameters for Burt's X^2 and the comparison X^2, respectively. Since Burt's X^2's are clearly nonsignificant, we can assume that $\lambda_0 = 0$. Hence, under the null hypothesis that $\lambda_1 = \lambda_0 = 0$, $(X^2_1/\nu_1)/(X^2_0/\nu_0)$ has an approximate F-distribution, $F(\nu_1, \nu_0)$, where ν denotes degrees of freedom. Comparing the X^2 for Burt's children with the Scottish boys, $F = 261$ ($p < 10^{-4}$) and with the Scottish girls, $F = 201$ ($p < 10^{-4}$).
63. See (*51*, item xiii), for details.
64. The data were taken from (*45*, table 189, p. 669). The sample was from groups I, II, and III of the white draft for World War I.
65. The details are given in (*54*, item i). The normal curve has sometimes been called the Quetelet curve [→ M. Kac, *Suppl. Adv. Probl.* **7**, 5 (1975)].
66. T. L. Kelley, *Fundamentals of Statistics* (Harvard Univ. Press, Cambridge, Mass., 1947), p. 278.
67. A. H. Halsey, *Br. J. Stat. Psychol.* **12**, 1 (1959).
68. Each new IQ category represented a nonoverlapping "ideal" occupational class. The category 141+ represented an ideal class I; category 127–141, an ideal class II; 115–127, an ideal class III; 103–115, an ideal class IV; 91–103, an ideal class V; and 50–91, an ideal class IV. Note that in the text there is a typographical error. The final border line was reported as 90, not 91 (*15*, p. 12). But in both tables III and IV the lowest category is given as 50–91. Moreover, if the category were 50–90, the column total would be 253 for the children and adults (from Burt's tables I and II), whereas Burt's column total is 261. Hence the lowest interval was evidently 50–91 for the reclassification given in Burt's tables III and IV.
69. Thus, Z^*_k is the inverse of $\sum_{i=1}^{k} P_{i\cdot}$. For example, if $\sum_{i=1}^{k} P_{i\cdot}$ were to equal 0.5, Z^*_k would equal 0. By definition, Z^{**} is the solution to $Z^{**} = (100 - X^{**})/15$, where X^{**} is Burt's category border line.
70. Since rounding to per mille can produce an error in a cell as large as 0.5, 141+ (category I) is always pooled with 127–141 (category II) in the forthcoming analyses.
71. Categories I and II are pooled.
72. W. Spielman and C. Burt, in F. Gaw, L. Ramsey, M. Smith, W. Spielman [under the general direction of C. Burt], in *A Study in Vocational Guidance, Report No. 33* (H. M. Stationery Office, London, 1926), pp. 12–17. The short section by Spielman and Burt was preliminary to the main investigation. In particular, "The main part of the research consisted of an intensive individual study of all the children due to leave the three selected schools [three elementary schools of a single London borough] in the course of the next twelve months. The number amounted to exactly one hundred—52 boys and 48 girls" (p. 10). The children were given general intelligence tests. Moreover, each child's home was visited and "an attempt was made to grade the intelligence of the mother—the parent most often interviewed. . . . Notes were made immediately on leaving the house, and the mother described as (1) Very Intelligent (A+); (2) Intelligent (A); (3) Moderately Intelligent (B); or (4) Unintelligent (C)" (p. 75). The general scheme was "drawn up by Dr. Burt" (p. 72). Clearly, the data from the main part of this investigation were not part of Burt's 1961 report. In searching for further details on Burt's 1961 study, McAskie and Clarke (*14*, p. 256) reported: "References to further details in other papers proved empty, and the trail through a number of publications led nowhere."
73. We assume that all vectors of six two-place percentages are equally likely and that the percentages sum to 100.
74. Burt provided no evidence whatsoever that his selected borough was representative of London as a whole for the research period of "nearly fifty years, namely from 1913 onwards" (*15*, p. 4).
75. C. Burt, *Br. J. Psychol.* **14**, 336 (1924). The percentages given in this paper are 0.1, 3, 12, 27, 36, 19, 3, and 0.2 going from vocational categories 1 to 8 (p. 349). Thus they are identical to the Spielman and Burt percentages on the three highest vocational categories. Spielman and Burt did not comment on the changed percentages for the lower vocational categories. Also, note the typographical error: Burt's 1924 paper is in volume 14, not 3.
76. Charles Booth's survey of London was done at the turn of the century [see C. Booth, *Life and Labour of the People in London* (Macmillan, London, 1892–1903), 17 vols.].
77. Spielman and Burt were probably referring to the 1921 English census, the last census before 1926. Perhaps both the 1911 and the 1921 London census figures were used, inasmuch as Booth's survey was evidently considered in the determination of the figures of Burt's 1924 paper (*75*). Recall that Burt's 1924 percentages and Spielman and Burt's 1926 percentages are identical for the three highest vocational categories.
78. Gottesman's statement of the duration of Burt's study is the most reasonable based on Burt's description: "The information was collected in the Greater London area during the period 1913 to 1960" (*24*, p. 36).
79. H. J. Eysenck, *Br. J. Math. Stat. Psychol.* **25**, i (1972).
80. I am particularly grateful to J. Simpson for assistance in the preparation of this article. I also thank J. O. Sines, I. Gormezano, J. A. Harvey, and W. L. Randall for valuable discussion; S. R. Rosner for some helpful comments; J. D. Broffitt, R. V. Hogg, and T. J. Robertson for generous statistical consultation; and M. Hacker for computer assistance.