

Sir Cyril Burt in Perspective

ARTHUR R. JENSEN *University of California, Berkeley*

ABSTRACT: *Numerical peculiarities in Sir Cyril Burt's reports of kinship correlations are most reasonably attributable to carelessness rather than fraud. Showing statistically significant deviations from randomness in the final digits of correlations, N s, and IQs, or other nonrandomness suggesting "digital preferences," is unconvincing as purported evidence that Burt's figures were faked. No analysis is given showing that Burt's results are biased so as to favor his theory, and they are in close agreement with numerous independent studies. Scientifically, in any case, the validity of Burt's theory of the polygenic inheritance of intelligence does not depend upon Burt's data.*

An indefatigable veteran of controversy throughout most of his long and eminent career, the late Sir Cyril Burt continues to provoke still new disputes more than 6 years after his death.

In the fall of 1976, sensational stories appeared in London newspapers under such banner headlines as "Crucial Data Was Faked by Eminent Psychologist" (Gillie, 1976) and "Theories of IQ Pioneer 'Completely Discredited'" (Devlin, 1976). In these and other articles, five psychologists (Michael McAskie, Alan and Ann Clarke, Leon Kamin, and Jack Tizard) are quoted as claiming that Burt's data on the inheritance of mental ability are fraudulent. Yet no real substantiation of the accusations accompanied these claims, either in the original articles or in the several letters to the *London Times* by McAskie and the Clarkes (November 13, 1976), Kamin (November 15, 1976), and Tizard (October 26, 1976) elaborating on their conjectures of fraud in Burt's data on IQ heritability.

Now, over a year later, McAskie (1978) serves up the first bit of seemingly intrinsic evidence for the indictment of Burt that supposedly amounts to more than what was admittedly just "claims and strong suspicions," and we are also promised future evidence "currently in preparation." The present method concocted by McAskie to detect fraud in Burt's figures was never even so much as hinted at in the newspaper articles, nor in the McAskie and Clarkes' letter to the *Times*, nor in my more re-

cent personal correspondence with McAskie's colleague, Ann Clarke (Note 1), concerning the basis of the charges against Burt.

May I suggest that we try to gain a proper perspective on the Burt affair.

At least a substantial first step in this direction was made in my 1974 article, which examined Burt's published kinship correlations that were the basis of his famous studies of the heritability of intelligence (Jensen, 1974). My original intention, in writing that article, was systematically to assemble, for the convenience of students of behavioral genetics, all the empirical results from Burt's many kinship studies that were scattered in various journals published from 1943 to 1972. Many of Burt's reports are cumulative, in the sense of enlarging sample sizes for certain kinship correlations and carrying over into later articles and reanalyses of IQ heritability some of the same correlations he had reported in previous papers. Also, some of the same errors in earlier articles get repeated identically one or more times in Burt's later articles. In the year-to-year cross-tabulations of all of this material in the nine large tables in my review, a number of peculiarities in some of the repeated correlations (r s) and their sample sizes (N s) became clearly evident. It seemed they must be erroneous, or at least certainly puzzling, and I clearly pointed this out, with the caveat that these questionable correlations must therefore be deemed useless for hypothesis testing involving genetic models of IQ variation.

What these errors, or at least peculiarities, in certain r s and their N s consist of can be easily summarized. Burt reported, altogether, 235 kinship correlations of various types (twins, siblings, parent-child, cousins, etc.), on three classes of variables: intelligence, scholastic achievement, and physical measurements. The sample size (N) is

Requests for reprints should be sent to Arthur R. Jensen, Institute of Human Learning, University of California, Berkeley, California 94720.

indicated for 165 *rs* and is not indicated, or not always explicitly, for 70 *rs*. Out of the 165 *rs* with explicit *Ns*, there are 20 instances where the *same r* (to three decimal places) is repeated in a later article accompanied by a *different N*. But these 20 invariant *rs* are not all independent, since often several correlations, for a number of mental tests and physical measurements, are based on the same sample. Thus the 20 invariant *rs* are attached to only 8 different *Ns*, out of a total of 48 different *Ns*. The errors, *prima facie*, would seem to be in the reported *Ns* rather than the *rs*. The $8/48 = 16.7\%$ "error rate" for *Ns* is not quite as high as the 20% rate of numerical errors that McAskie finds in a large sample of journal references from Burt's articles. We cannot appraise how abnormally high this error rate is without comparative studies, but one can hardly imagine that these reference errors were motivated by Burt's theoretical position on the inheritance of mental ability. So many numerical errors, however, do seem surprising in the context of the very sophisticated genetical and statistical treatment and the virtual absence of any missteps at the conceptual and theoretical level.

What else can we note about these changed *Ns* with the invariant *rs*? In four cases, the first reported *N* is smaller than the second reported *N* for a given set of invariant *rs*, and in four cases we see the reverse. It seems strange that four of Burt's sample sizes for certain kinships should *decrease* from one report to the next. Three of the eight *N* changes involve a reversal of digits or a substitution of one different digit from the first to the second *N*; the remaining are entirely different *Ns*.

In preparing my original (1974) analysis of Burt's correlations and their peculiarities, summarized above, I was concerned not with Burt's character but with the scientifically more intrinsic questions of whether the peculiarities were inconsistent, careless errors, or showed any consistent slant favoring his polygenic theory of the inheritance of intelligence, or were in any way significantly discrepant from comparable results of other scientifically acceptable studies. I found no evidence of such theoretical bias in Burt's errors, and I later (Jensen, 1976) offered as the most parsimonious explanation sheer carelessness on Burt's part, however damaging that interpretation certainly must be to his scientific reputation, and however incongruous it may appear in light of his superb technical command of psychometrics, statistics, and quantitative genetics. I confess I am

deeply puzzled by it. But my own analyses have not revealed the directional biases in Burt's errors that one should expect in them if they were intentionally slanted to favor his genetic theory of mental ability. For example, the invariant *rs* occur with almost equal frequency in the measurements of intelligence (6 cases), scholastic achievement (7 cases), and physical characteristics (7 cases). I statistically compared the frequency distribution of IQ differences between Burt's 53 pairs of identical twins reared apart with the composite distribution of similar data from three other studies reported by independent investigators, by means of the Kolmogorov-Smirnov two-sample test, which is designed to detect differences in all of the moments of two frequency distributions, and it showed no significant difference between Burt's distribution and the other three (Jensen, 1974, pp. 15-16).

In the same vein, Rimland and Munsinger (1977) looked at Burt's correlations for various kinships in relation to practically all of the correlations for the same kinships reported in 42 independent studies involving over 30,000 correlational pairings from 8 countries in 4 continents and originally summarized by Erlenmeyer-Kimling and Jarvik (1963). Rimland and Munsinger (1977, p. 248) point out that "the deletion of Burt's data would have no appreciable effect on the overall picture . . . Burt's figures differ [with an average deviation of .03] from the median values of the many authors in an unsystematic way," as shown in Figure 1. Thus, McAskie's remark that I (or anyone else) had been "misled" by Burt is pure nonsense, if by "misled" it is implied that Burt's findings and conclusions yield a picture that is significantly at odds with the scientific consensus provided by the numerous other studies of IQ heritability, both prior and subsequent to Burt's own publications.

Another point consistent with the "carelessness" explanation is shown in Table 2 (p. 11) of my 1974 article, which reproduces Burt's tabulated *rs* from the well-known monograph on twins by Newman, Freeman, and Holzinger (1937): Alongside of Burt's listing I put the *correct rs* from Newman et al. There is about the same rate of errors or peculiarities in Burt's listing of the Newman et al. *rs* as for his own data. Some of these are transparently due to miscopying *rs* from the wrong row or column of a table in Newman et al. (e.g., *rs* for head length and breadth are interchanged). There seems to be no sense to these several errors and Burt surely realized they could be checked against

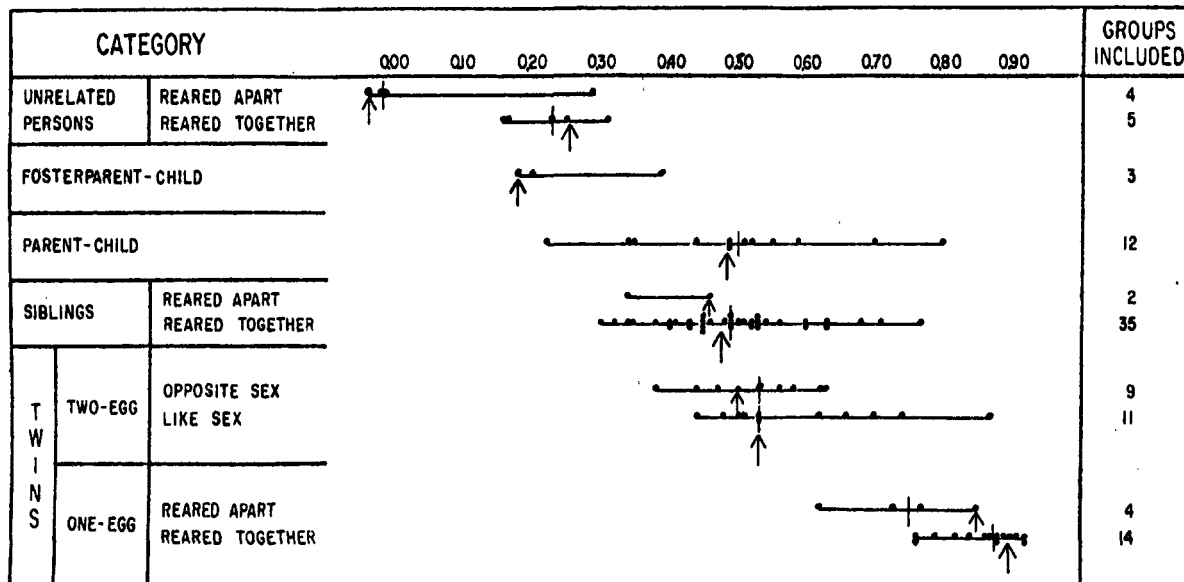


Figure 1. Correlation coefficients for "intelligence" test scores from 52 studies. Some studies reported data for more than one relationship category; some included more than one sample per category, giving a total of 99 groups. Over two-thirds of the correlation coefficients were derived from IQs, the remainder from special tests (for example, Primary Mental Abilities). The midparent-child correlation was used when available; otherwise, the mother-child correlation. Correlation coefficients obtained in each study are indicated by dark circles; medians are shown by vertical lines intersecting the horizontal lines which represent the ranges. (From "Burt's IQ Data," letter by B. Rimland and H. Munsinger, *Science*, 1977, 195 (January 12, 1977), 248. Copyright 1977 by the American Association for the Advancement of Science. Reprinted by permission.)

the original tables in Newman et al. This hardly looks like fraud.

Since the appearance of my 1974 article, the claims of Burt's critics have escalated from the initial prima facie evidence of carelessness, to charges of bias, and now, finally, to the worst crime a scientist can be accused of—publishing purportedly empirical data and results that were only manufactured out of whole cloth to promote their author's own theoretical position—in short, absolute fraud.

It seems to me extremely difficult, although of course never wholly impossible, to imagine that a scholar of Burt's distinction, phenomenal industry, and pioneer dedication to the development of psychology as a quantitative, scientific discipline could actually be guilty of such a charge. It would indeed be disillusioning, but, after all, it is now believed that Gregor Mendel doctored his data (Fisher, 1936) and Isaac Newton fudged his figures (Westfall, 1973) to fit their theories, so really no one can ever be regarded as entirely above suspicion. But we do know that Burt was an inveterate statistical analyst, and he had ready access to

enormous sources of test data during his 20-years' tenure as chief psychologist to the London schools. So the idea that he would have invented all these data on twins and various other kinship correlations seems bizarre, to say the least. Moreover, a well-known British researcher on twins, James Shields, informed me in personal correspondence, long before the recent charges of fraud, that in the course of his own research he came across a number of twins who said they had been tested by Burt. Also, Burt's biographer, Leslie Hearnshaw, the leading historian of British psychology, has informed me of independent testimony of a twin who was tested by Burt's assistant, Margaret Howard. (Burt's detractors originally claimed that Miss Howard was a fictitious person invented by Burt as a co-author of one of his important articles, but she has since been identified as a real person by a psychology professor at Manchester University, John Cohen, who testifies he knew Miss Howard while he was a PhD student under Burt [Cohen, 1976].)

Neither does the notion of fakery accord with the testimony of a distinguished colleague of Burt's,

the geneticist J. A. Fraser Roberts, FRS, who writes: "Much of our work overlapped with that of . . . Burt, with whom I was closely associated. I had many long personal meetings with him, during which we went through his data and ours. I found him thoroughly accurate and reliable and our results were in close accord. I should like to condemn most strongly the idea that he cooked his data" (Roberts, 1976). This testimony jibes with my personal impressions of Burt, gained during my many memorable visits to his London flat during his last years, which I have described elsewhere (Jensen, 1972).

McAskie's method for attacking Burt's integrity, by showing nonrandom distributions of terminal digits and other "digital preferences," is not only half-baked, as presented, but is, in principle, incapable of standing up as evidence of fraud in Burt's (or in anyone else's) data. In the first place, one can always find in any limited set of random numbers a few statistically nonrandom features. There are numerous possible "digital preferences" that can be demonstrated when persons are required to make up numbers at random, and various examples of these can be found at a statistically significant level in a finite set of digits from a table of random numbers. We are given no idea of how many possible kinds of digital nonrandomness McAskie considered and tested for significance. For example, another type of preference is for *odd* versus *even* digits. But this particular preference happens not to show up significantly in Burt's terminal digits, either for *rs*, *Ns*, or *IQs*. Another preference is for central digits (i.e., 4, 5, 6) versus extremes (i.e., 1, 2, 3, and 7, 8, 9); and this too is nonsignificant in Burt's data. We could go on and on with this game and be assured of finding, somewhere, statistically significant departures from randomness. This fact is the well-known reason for most scientists' skepticism concerning the seemingly impressive statistical evidence for ESP: We know all too little of the whole population of possible outcomes from which we are shown only a few instances of statistically significant outcomes.

But McAskie seems quite unaware of the fact that this type of analysis, even if carried out to the limit and reported much more thoroughly than he has attempted, is, in any case, simply futile as evidence for the main point he wishes to make. Statisticians have known for many years that people typically produce nonrandom distributions of final digits or of digit repetitions when they copy num-

bers, even when the data recorders are careful, well-trained, and have no axe to grind. This is true when numbers are read off scales, such as calipers, rulers, dials, and slide rules, as well as for digital displays. Many examples of this are provided in *The Advanced Theory of Statistics* by Kendall (1948, pp. 187-191), who notes that "even those who are aware of the existence of the possibility of bias and the necessity for taking great care . . . may nevertheless fail to avoid it" (p. 189). And, "It is abundantly clear that we must look for true randomness elsewhere than in the mere lack of purpose on the part of human observers" (p. 190). This, of course, is why in this day of computer analysis, we routinely verify, and even, at times, double-verify, the keypunching of data cards. Burt used only a desk calculator and slide rule for his computations.

Thus it seems evident that the nefarious deeds of which McAskie accuses Burt are in reality the result of a human frailty found universally in all of us, no doubt including McAskie himself. I trust that we shall not hear from McAskie again on this matter until he has carefully reviewed his own work and can assure us that he himself is more immune to digital preferences and is a more accurate recorder of final digits than was Burt.

In any case, Burt's theory of the inheritance of intelligence does not depend now upon his own data, nor can the truth of the theory be at all affected by any new revelations as to Burt's personal character, for better or worse. That is a matter, not of scientific, but of purely historical and biographical interest. Biographical research is, I believe, the only possibility, at this point, of assessing the likelihood that Burt made up fictitious data. For example, there does remain some mystery as to just where, when, and how Burt obtained the test data on the rare 32 identical twins reared apart who were added to his previous twin data between the years 1955 and 1966, when Burt was 73 to 83 years of age. Burt's biographer, Leslie Hearnshaw, is fully aware of these questions and, I trust, will leave no avenue unexplored for getting at the relevant facts. His excellent credentials as an objective and impartial biographer and historian of psychology lead me to recommend that we wait for his forthcoming judgment on Burt's career.

Meanwhile, I think we must take a skeptical view of any of Burt's critics to whom Burt's real sin was not that his data were questionable, but that he believed in a general factor of mental ability which

is largely inherited. Burt's critics obscure the necessary distinction between biography and science. If their interest is in the former, why are they not doing biographical research? If their interest is in the latter, why are they not conducting their own investigations on the genetics of intelligence to determine if Burt's findings can be replicated? That is the way of science.

REFERENCE NOTE

1. Clarke, A. Personal communication, January 31, 1977.

REFERENCES

- Cohen, J. Heredity and intelligence. Letter to the *London Times*, November 10, 1976.
- Devlin, T. Theories of IQ pioneer "completely discredited." *London Times*, October 25, 1976.
- Erlenmeyer-Kimling, L., & Jarvik, L. F. Genetics and intelligence: A review. *Science*, 1963, 142, 1477-1479.
- Fisher, R. A. Has Mendel's work been rediscovered? *Annals of Science*, 1936, 1, 115.
- Gillie, O. Crucial data was faked by eminent psychologist. *London Sunday Times*, October 24, 1976, pp. 1-2.
- Jensen, A. R. Sir Cyril Burt (obituary). *Psychometrika*, 1972, 37, 115-117.
- Jensen, A. R. Kinship correlations reported by Sir Cyril Burt. *Behavior Genetics*, 1974, 4, 1-28.
- Jensen, A. R. Heredity and intelligence: Sir Cyril Burt's findings. Letter to the *London Times*, December 9, 1976, p. 11. (Reprinted in *Phi Delta Kappan*, 1977, 6, 471; 492)
- Kamin, L. J. Heredity and intelligence: Research in perspective. Letter to the *London Times*, November 15, 1976.
- Kendall, M. G. *The advanced theory of statistics* (Vol. 1; 4th ed.). London: Griffin, 1948.
- McAskie, M. Carelessness or fraud in Sir Cyril Burt's kinship data? A critique of Jensen's analysis. *American Psychologist*, 1978, 33, 496-498.
- McAskie, M., Clarke, A., & Clarke, A. D. B. Heredity and intelligence: Sir Cyril Burt's reputation. Letter to the *London Times*, November 13, 1976.
- Newman, H. H., Freeman, F. N., & Holzinger, K. J. *Twins: A study of heredity and environment*. Chicago: University of Chicago Press, 1937.
- Rimland, B., & Munsinger, H. Burt's IQ data. *Science*, 1977, 195, 248.
- Roberts, J. A. F. Studies in heredity. Letter to the *London Times*, November 29, 1976.
- Tizard, J. Heredity and intelligence. Letter to the *London Times*, October 26, 1976.
- Westfall, R. S. Newton and the fudge factor. *Science*, 1973, 179, 751-758.