IQ and Heredity: Suspicion of Fraud Beclouds Classic Experiment

Charges of scientific fraud, as yet unproved, have been made against an eminent English psychologist, the late Cyril Burt, whose work has featured prominently in the debate about racial differences and intelligence.

Burt is accused by his critics of having doctored or even invented his extensive and partly unique collection of IQ test data in order to support his theory that intelligence is determined primarily by heredity.

Should the accusation prove true, the forgery may rank with that of the Piltdown Man in that for years it remained undetected while occupying a pivotal place in a fierce scientific controversy.

Burt's work has been used not only in the United States, notably by hereditarians such as Arthur Jensen of the University of California and Richard Herrnstein of Harvard, but also in England where Burt wielded considerable influence over national educational policy. As a government adviser in the 1930's and 1940's, he was influential in setting up the three-tier system of British education. In accordance with Burt's views that intelligence is largely innate, children were irredeemably assigned to one of the three educational levels on the basis of a test given at the age of 11.

Burt's conclusions have been under suspicion for several years because of internal inconsistencies in his data. What has sparked off accusations of outright fraud is an article in the London *Sunday Times* of 24 October. The newspaper's medical correspondent, Oliver Gillie, reported his failure to find any evidence that Burt's two chief coauthors on his later papers, Miss Margaret Howard and Miss J. Conway, had ever existed.

Gillie could find no sign of Howard or Conway in the records of London University, the address given on their scientific papers, and none of 18 acquaintances of Burt could remember having ever met or heard of them. "It must be considered a possibility," Gillie concluded, "that Margaret Howard and J. Conway never existed, but were the fantasy of an aging professor who became increasingly lonely and deaf." Burt died in 1971 at the age of 88, having retired from the chair of psychology at University College, London, some 20 years earlier.

For those who were in any case skeptical of the notable fit between Burt's data and his theories, the failure to trace his collaborators has come as clinching evidence of fraud. Leon Kamin of Princeton University says of Burt's work, "It was a fraud linked to policy from the word go. The data were cooked in order for him to arrive at the conclusions he wanted." On the other side of the Atlantic Liam Hudson, professor of psychology at Edinburgh University, considers that the inconsistencies in Burt's data and the difficulty in tracking down his coauthors put the question of Burt's fraudulence "beyond argument."

Other psychologists, however, believe that an innocent explanation, such as misprints or carelessness, may account for the peculiarities in Burt's data. As for his missing coauthors, the *Sunday Times* reported 2 weeks after its first article on Burt that Miss Howard was at least not entirely a figment of Burt's imagination. In the 1930's she was, despite the newspaper's failure to find her in the records, a faculty member of London University.

Burt's continuing importance has been due to the still unrivaled collection of data on heredity and IQ which he gathered as research psychologist to the London school system from 1913 to 1932, and later as professor of psychology at University College, London. Burt's data were cited by Jensen in his furiously debated *Harvard Educational Review* article of 1969. The article suggested that the large genetic component of intelligence might underlie racial differences in IQ scores and hence the failure of compensatory education programs addressed to minorities.

Herrnstein selected Burt's data specifically to support his almost equally provocative argument (*The Atlantic*, 1971) that social standing is partly based on inherited differences in intelligence. Burt's array of kinship and IQ data is not crucial to the hereditarians' arguments, but it has probably played as important a role in their case as has any other collection of data.

It was because of its central role in the debates touched off by the Jensen and Herrnstein arguments that Burt's work—apparently for the first time—came under serious critical review. The critic was Kamin, a psychologist who specializes in the conditioned reflex and who had never ventured into the IQ field until 1972 when a student urged him to read one of Burt's papers. "The immediate conclusion I came to after 10 minutes of reading was that Burt was a fraud," he says. Being an outsider to the field, Kamin spotted what no one inside had seen, that Burt's results are riddled with internal implausibilities and basic methodological oversights.

For example, the pearl of Burt's data collection is a survey of separated identical twins. Since the twins have the same genes, any difference in their intelligence should be due to environment alone, so that the correlation of their IQ scores in theory gives a pure measure of the influence of heredity on intelligence. This is a classical experiment in psychology, but one that is not often done because of the rarity of separated identical twins.

Burt published the first full report of his twins in 1955, when he had located 21 pairs, a second report in 1958 when the collection amounted to "over 30" pairs, and a final report in 1966 with 53 pairs. What Kamin noticed was that the correlation between the IQ scores of the separated twins was given as 0.771 for all three studies. For the correlation to remain unchanged through different sample sizes is improbable even at one occurrence, but the IQ correlation between identical twins reared together sticks at 0.944 through three different sample sizes, and there are many other such examples. For whatever reason, there is a strange imperturbability in the results Burt obtains from a changing data base.

Kamin also noticed that Burt often failed to record facts quite basic to the methodology of his surveys, such as the sexes of the children, the age at which they were tested, or even what particular test was applied. "The numbers left behind by Professor Burt," Kamin concluded in a lecture given in 1973, "are simply not worthy of serious scientific attention."

Kamin's views on IQ lie at the environmentalist pole of the debate, his position being that there is no way of proving that intelligence is inheritable. His opinions of Burt are not universally accepted. The suggestion of fraud "is so outrageous I find it hard to stay in my chair," said Herrnstein when asked his opinion of Kamin's criticisms. "Burt was a towering figure of 20th century psychology. I think it is a crime to cast such doubt over

a man's career." But Burt's empirical legacy has probably been harmed beyond repair, and the man who delivered the coup de grace is none other than Jensen.

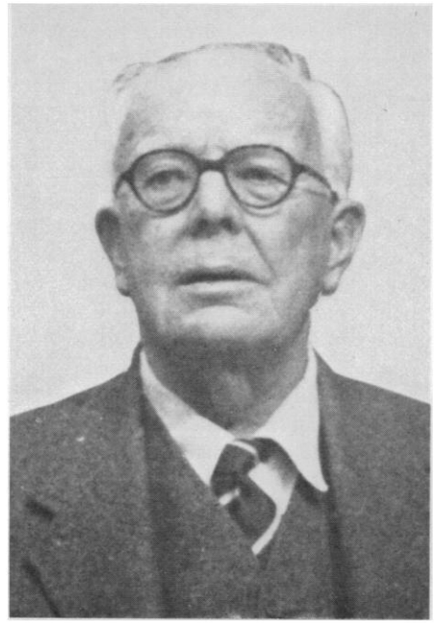
Jensen greatly admired Burt, came to know him in the last years of his life, and described Burt in an obituary as "one of the world's great psychologists." As a service to other psychologists, Jensen decided to make a compendium of all the data on kinship correlations which Burt had published in scattered form in numerous papers. On a visit to England in May 1972 he gathered a full set of reprints from Burt's former housekeeper and started to tabulate all Burt's data. In so doing he came across the same curious consistencies noted by Kamin.

Any particular instance of an invariant correlation with a changing sample size, Jensen observed, "can be rationalized as being not too improbable. But 20 such instances unduly strain the laws of chance and can only mean error, at least in some of the cases." Burt's correlations "are useless for hypothesis testing," he concluded in an article which appeared in *Behavior Genetics* in 1974. As to the reason for the errors, Jensen speculated, "It is almost as if Burt regarded the actual data as merely an incidental backdrop for the illustration of the theoretical issues in quantitative genetics, which, to him, seemed always to hold the center of the stage."

It is greatly to Jensen's credit that he was prepared to read out of the scientific literature, as his article essentially does, data which had been an important buttress of his own controversial position. Moreover, Burt was a man to whom Jensen was linked both by belief and academic lineage (Jensen was a postgraduate student of H. J. Eysenck, who was a pupil of Burt's). As to the credit for discovering Burt's errors, Jensen's article was the first to be published (Kamin's appeared in his book *The Science and Politics of IQ*, also 1974), but Kamin was the first to comment on the errors, in lectures given around the country in 1972 and 1973. Jensen did not personally attend any of these lectures but acknowledges in a footnote that Kamin was the first to draw attention to the invariant correlation of the twin studies.

Even Eysenck, whom Burt considered his brightest pupil, has conceded that Burt's data are unusable. The problem described by Jensen, Eysenck said in a recent letter to the London *Sunday Times*, "makes it impossible to rely on these figures in the future."

Are the errors in Burt's papers the results of a deliberate attempt to deceive, or simply the inattention to detail of an elderly and ailing man? Burt is not around to defend himself, but both Eysenck and Jensen believe that there are innocent explanations of what happened. Kamin, on the other hand, has come to suspect that Burt consistently invented data from his very first work published in 1909.



Cyril Burt, as photographed in 1971 by Arthur Jensen. From an obituary by Jensen in Psychometrica *June 1972.*

One event that has made the issue harder to resolve is the destruction of Burt's original data. After his death his housekeeper invited colleagues to his apartment to take their pick of the books and papers she could not place with libraries. Burt had half a dozen tea-chests full of papers and the housekeeper asked advice about them of two people who happened to be visiting the apartment, one of whom was Liam Hudson of Edinburgh University. Hudson, who is not one of Burt's academic allies, says that the tea-chests were full of raw test sheets, unfiled and impossible for anyone but Burt to make sense of. He told the housekeeper, he says, that she would be "perfectly justified in burning them." Asked if the tea-chests might not have contained data bearing on the disputed twin studies, Hudson replies, "What I was doing was walking around a gloomy house with a distraught housekeeper. Whether they contained anything of genuine significance I just don't know."

Most of Burt's data on IQ and kinship correlations were published after his retirement in 1950. According to Leslie Hearnshaw of Liverpool University, who is preparing a biography, Burt was hard up during this time and had to do a lot of hack work to make ends meet. He was suffering from a form of deafness known as Ménière's disease which made it hard for him to communicate. He wrote several papers on psychical research, in which he had long been interested. Philip Vernon, a collaborator of Burt's who is now at the University of Calgary, Alberta, says of Burt's state of mind that he was always a bit paranoiac. "He would train students brilliantly but once they turned against him he would turn on them with a vengeance. He was very helpful if one was with him, as long as one did not criticize the theories he built up in the 1910's and 1920's. Unfortunately, he never changed his theories."

Old, ailing, embattled in numerous controversies, it is quite possible that Burt was simply careless with his data and neglectful of accuracy. Jensen, for example, says that though he can't offer any explanation for the errors, "they seem to be too haphazard and not planned. If Burt was trying to fake the data a person with his statistical skills would have done a better job."

Eysenck's suggestion, somewhat more damaging to Burt, is that to avoid the chore of recalculating the correlations with the new data, Burt simply carried over the correlation figures from his earlier papers. Though quite inadmissible as a scientific procedure, a shortcut of this nature would fall a long way short of fraud since not done with intent to deceive.

A similar explanation is proposed by G. C. Drew, present occupant of Burt's chair of psychology. "As he got old he was remembering old figures that stuck in his mind from earlier papers." But Drew still sees the affair as purely a matter of carelessness. "Burt was totally convinced of the rightness of his views and he became exceedingly careless with the data," he says.

Such evidence as suggests that Burt might have been inventing data revolves chiefly around Burt's two collaborators, Miss Margaret Howard and Miss J. Conway. The failure of the London *Sunday Times* to find them or anyone who knew them is provocative but not conclusive. Indeed the paper has since reported the account of a Manchester University professor who knew Howard at least in the late 1930's: "She used to wear tinted glasses and a dark blouse with a ribbon. She had a lovely smile," the professor remembers.

Even if Howard and Conway existed once, there is another reason for doubting their reported collaboration with Burt in his crucial papers of the late

1950's. Howard and Conway appear frequently during this period as the authors of book reviews in the journal that Burt edited, the *British Journal of Statistical Psychology*. The style of the reviews is similar to Burt's own highly distinctive style. It is of course possible that the similarity reflects Burt's hand as editor, not as author. But the reviews struck people even at the time as being Burt's work. Vernon, for example, says he had long suspected that Burt used pseudonyms for the book reviews, an act that he describes as "silly but not particularly heinous."

Use of Howard and Conway as pseudonymous book reviewers may indeed have been harmless enough, and in adding their names as coauthors to papers in the 1950's Burt perhaps meant only to give credit for research they had indeed done much earlier, maybe in the 1930's. Even articles of which Howard or Conway appear as the sole authors are, on the pseudonym hypothesis, evidence more of eccentricity than intent to deceive. What is much harder to explain is the fact that Burt was almost certainly incapable of administering IQ tests himself after 1950 and would have required real collaborators to test the twins from whom new results are reported in the articles of 1958 and 1966. In fact Burt wrote in 1969 to a correspondent that when travel was necessary to interview a twin, "The job was delegated to Miss Conway or Miss Howard." If Howard and Conway were by that time only pseudonyms, who tested the twins?

Reading Burt's final paper on the twin survey in the light of such suspicions, it is hard not to be struck by the patness with which crucial, yet hitherto unpublished, data are adduced to demolish each of his critics' arguments. The paper appeared in the *British Journal of Psychology* in 1966, when Burt was 83. The number of separated identical twin pairs had grown from "over 30" in 1958 to 53, making the study by far the largest of its kind. But the study was now revealed to possess a feature that made it even more unique and authoritative. The anti-hereditarians had been claiming that the high IQ correlations found between separated twins might have a lot to do with a correlation between their environments. This was a plausible argument—and one that is true of all the other twin studies—because adopted children are indeed usually placed in homes similar to their own. Though it was not mentioned in the earlier reports of 1955 and 1958, the 1966 paper now reveals the remarkable fact that the homes of Burt's twin pairs, as judged by the occupational categories of the parents and foster parents, are entirely uncorrelated. (The correlation, though Burt gives only the raw figures, is − 0.4.) "These figures," Burt then proceeds to observe, "should dispose of one of the commonest explanations advanced by thoroughgoing environmentalists—namely, that the high correlations for the separated twins is due to the way the foster-parents were chosen."

Burt's figures were unlikely enough to have prompted an inquiry from at least one psychologist. Sandra Scarr-Salapatek of the University of California wrote in 1970 asking Burt for more information because the data "looked funny" to her. And Vernon says he thought at the time that the result was highly unlikely. "I could not stomach that, I could not believe that. I didn't know what he had done." Asked why no one had disputed Burt's result, Vernon says that "there were certainly grave doubts although nobody dared to put them into print, because Burt was enormously powerful." Burt's power seemed to have stemmed not so much from patronage—he was by then retired—as from the way he would use his formidable prose style and mastery of statistics to take out after his critics. "He would write a 50-page paper denouncing any criticisms," says Vernon.

Hearnshaw, Burt's biographer, believes that, for whatever reason, it is only Burt's later work that is flawed. Kamin, on the other hand, is coming increasingly to suspect that Burt concocted data from the very beginning of his research career. "He was doing it very flagrantly at the end of his life. Earlier he was more careful in covering his tracks," says Kamin. This intriguing hypothesis is based partly on what Kamin sees as a succession of implausible results which accord with Burt's apparent social biases, and partly on a "pattern of consistent obscurity" that runs through Burt's papers; he rarely gives primary data but refers the reader to unpublished reports, at least some of which Kamin has found to be as elusive as Howard and Conway.

These two authors, Kamin notes, appear frequently in the 1950's as contributors to Burt's journal, writing in a style that is "unmistakably Burtian," but their literary endeavors ceased almost immediately after Burt quit as editor. They and other reviewers with a similar style claim priority for Burt in various discoveries, demolish books written by his critics, and heap praise upon books written by Burt. "It is inconceivable that the writings of Howard and Conway should have been by anyone but Burt," Kamin says. He rejects the idea that the pseudonyms and other flaws can be attributed to senility. The twins, for example, many of whom Kamin suspects may never have existed, are referred to in passing in a paper of 1943 when Burt was at the height of his powers. "I suspect that everything the man did from 1909 is wholly fraudulent," Kamin suggests.

The flaws in Burt's work, whatever the reason for them, are obvious enough now that Kamin and Jensen have pointed them out. Yet Burt's work was never challenged during his lifetime. He was preeminent among his colleagues in England, being the first psychologist to receive a knighthood, and the American Psychological Association awarded him its Thorndike prize in 1971. At least from 1969 onward, his data occupied a central position in controversy, in a subject which is presumably no less rigorous than other disciplines. Why were the flaws not detected earlier? Why did journal editors and journal boards not require that he report his results in a more complete or accurate form?

"The sober fact is that scholarly penetration of the literature, and endless delving into primary sources, occurs only very rarely," suggests Hudson. "It reflects on us all that these figures should have been in the literature of a highly contentious and important area for more than a decade before anyone went back to examine them as Kamin did. It strikes me as very damaging to us as a profession that articles were coming from someone called Conway whom no one had ever heard of. That is not the way that a community of scholars should be working."

Kamin's interpretation is that Burt's data remained unchallenged because they confirmed what everyone wanted to believe. "Every professor knew that his child was brighter than the ditch-digger's child, so what was there to challenge?" The moral of the tale, according to Kamin, is "Caveat emptor! The people who buy social science should remember that those who have collected the data may have axes to grind."

Others see the episode as less far-reaching. Scarr-Salapatek, one of the early doubters of Burt's data, says that "people trusted Burt to be reporting accurately what he did, so I don't think it is surprising they accepted his data even if they were implausible." In fact, apart from the strange lack of environmental correlation in the twin study, Burt's results for the IQ correlations of his twins are well in line with those of other studies. This, and the fact that the data are published in such scattered form that the discrepancies are not readily apparent, are reasons why the flaws remained undetected, suggests Jensen.

As for the practical implication of removing Burt's data from the scientific literature, there is, naturally, a range of opinion. Kamin states that the data have played a "monumentally important role," particularly those to do with certain rare kinship correlations (such as between second cousins). Richard Lewontin of Harvard, an anti-hereditarian and eminent population geneticist, says that Burt's twin study "is the only large study which is methodologically correct, so its loss is no trivial problem for the heritability people. It is also not nice for them to have this mess in their backyard," he observes. But according to Jensen, all Burt's salient results have now been duplicated, and no one was led to any conclusions which they would not have reached even if Burt had never existed. "It is not like the Piltdown man which led people into error," Jensen says, adding that he does not mean to imply by the comparison that Burt's data too were forged. Scarr-Salapatek, on the other hand, thinks that subtraction of Burt's data "results in a downward estimate of heritability, though not a radical reassessment." The reassessment might require the hereditarians to reduce their estimate of the heritability of intelligence from 80 percent (Herrnstein, 1971) to nearer 60 percent, the figure arrived at by Christopher Jencks of Harvard without using Burt's data.

Burt's data are probably now unusable in any case, but 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. The only sure evidence of error, the invariant correlations, is a curious mistake for a cunning forger to make. Perhaps, when old and ill, Burt was too proud to ask for help in doing the calculations and, as Eysenck suggests, carried over the results from earlier papers. That is probably the most plausible present explanation for those who like economy in hypotheses. The question is how well it stands up under such burdens to belief as the failure to locate Misses Howard and Conway, the implausibility of some of Burt's results, the apparent use of pseudonyms, and any other suspicions that may accumulate. Should it not, Burt would have much to answer for.—NICHOLAS WADE

Superport for Palau Debated: Ecopolitics in the Far Pacific

Although the planners insist that it is still only a "concept," the possibility that a superport for transshipping Iranian oil to Japan will be built in a remote island group in the western Pacific has spurred American environmental groups to concerted opposition.

The proposed location for a deepwater port, where supertankers from the Middle East would deliver oil for transfer to smaller tankers, is the island group of Palau in the western Caroline Islands, part of the U.N. Trust Territory of Micronesia administered by the United States.

Initially, only port and oil storage facilities would be constructed, but the proponents of the plan contemplate later addition of a refinery and energy-related industry to form what they call an "energy-industry complex."

Backing the plan are Japanese industrial and financial interests. The Iranian government also apparently would be a partner, and Iran would be involved both as a supplier of oil and investor. A pivotal role in the formative stages of the project has been played by Robert Panero, a New York consultant specializing in resource development. Panero has orchestrated the complicated negotiations, now advanced to the point where a feasibility study is to be carried out.

Opposition to the project has come from some Palauans—the islanders appear to be divided on the issue—and from American scientists and environmental organizations. Scientists have been attracted to the islands by the particularly rich and fragile ecology of Palau's coral reefs and enclosed lagoons. They fear the effects of dredging and construction on the ecology of the islands, and also predict that the economic development and increased population which would accompany a half-billion-dollar project would have a devastating effect on the Palauan culture.

This concern was expressed in a resolution passed at the Pacific Science Congress in Vancouver last year, which made the point, among others, that, "marine scientists who have worked on Palau and in other areas of the Pacific consider the site of the proposed development to be of value (scientific and park) unequalled in Oceania and of an order of importance rendering it eligible for designation as a World Heritage Area as defined by UNESCO." The resolution "strongly urges the appropriate authorities that this project should be abandoned on scientific grounds and because of the potential adverse effects upon the human population and biota of Palau."

Environmentalists in the United States, Japan, and elsewhere have taken up the issue. In this country organizations such as the Audubon Society, Sierra Club, and Environmental Defense Fund have shown interest, with the Natural Resources Defense Council apparently seeking to organize a coalition effort.

Palau is the westernmost territory now under direct U.S. control. An archipelago curving some 250 kilometers north and south on the edge of Micronesia, the islands have about 14,000 inhabitants, Palau's colonial history began with its discovery by the Spanish in 1543 and annexation a century later. In 1899 Spain sold the islands to Germany, and after World War I they were put under League-of-Nations mandate and administered by the Japanese. Palau was an objective in the American island-hopping offensives of World War II and briefly made the news in 1944 when U.S. Marines took Palau's Pelelieu Island where the Japanese had an air strip. After the war, Palau was part of the U.N. Trust Territory assigned to the U.S. The trust ends in 1981 and Micronesia, of which Palau is a district, is scheduled to vote on independence or association with the United States. The Palauans are said to be most likely to opt for self government, but to retain a tie with the United States, which would handle defense and foreign policy affairs.

The decision is expected to be crucial to the life of the project because the Japanese promoters of the superport are said to stress political stability and to believe that only association with the United States would assure military protection and stable government.

The idea of the superport is in part a product of the change in the economics and technology of the oil trade. The main