

THE MISMEASURE OF MAN

Stephen Jay Gould

'If the misery of the poor be caused not by the laws of nature, but by our institutions, great is our sin' - Charles Darwin, *Voyage of the Beagle*

REVISED AND EXPANDED EDITION



PENGUIN BOOKS

PENGUIN SCIENCE

CAN HUMAN INTELLIGENCE BE MEASURED?

In this highly acclaimed account, Stephen Jay Gould effectively discredits both the assertion that intelligence can be measured as a single, innate heritable quantity and its resulting conclusions, such as the idea that oppressed and disadvantaged groups – races, classes, or sexes – are innately inferior and deserve their status. Tracing the history of scientists' attempts to assess human intelligence from nineteenth-century craniometry to today's far more sophisticated methods of IQ testing, he exposes the fallacies of biological determinism.

Now revised and expanded, this edition includes an extensive new introduction and additional essays. As well as answering questions on race, racism, and biological determinism in general, he issues a direct response to a recent manifestation of the theory of innate intelligence, *The Bell Curve* (1994, Herrnstein and Murray), and shows that there is nothing new in this latest attempt to present social inequalities as dictates of biology.

Elegantly written and tightly argued, *The Mismeasure of Man* provides a brilliant riposte to those who would classify people according to their supposed genetic inheritance and reaffirms the Darwinian belief in the variety of human potential.

'This is a marvellous book, and it is marvellous in lots of different ways ... The history of these attempts to measure the intellectual and moral worth of individuals ... turns out to be a detective story to rival most of Agatha Christie's'

– *Sunday Times*

'Superlative historical scholarship and critical analysis' – *Nature*

SECOND EDITION

Cover illustration by Chris Riddell



PENGUIN
Science
Psychology

ISBN 0-14-025824-8



9 0 3 0 1



PENGUIN BOOKS

Published by the Penguin Group

Penguin Books Ltd, 27 Wrights Lane, London W8 5TZ, England

Penguin Books USA Inc., 375 Hudson Street, New York, New York 10014, USA

Penguin Books Australia Ltd, Ringwood, Victoria, Australia

Penguin Books Canada Ltd, 10 Alcorn Avenue, Toronto, Ontario, Canada M4V 3B2

Penguin Books (NZ) Ltd, 182-190 Wairau Road, Auckland 10, New Zealand

Penguin Books Ltd, Registered Offices: Harmondsworth, Middlesex, England

First published in the USA by W. W. Norton & Co., Inc. 1981

Published Simultaneously in Canada by George J. McCleod Ltd

Published in Pelican Books 1984

Reprinted in Penguin Books 1992

Revised and expanded edition first published in the USA by W. W. Norton & Co., Inc. 1996

First published in Great Britain in Penguin Books 1997

10 9 8 7 6 5 4 3

Copyright © Stephen Jay Gould, 1981, 1996

All rights reserved

Printed in England by Clays Ltd, St Ives plc

Set in Baskerville

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser

Contents

Acknowledgments	15
Introduction to the Revised and Expanded Edition: <i>Thoughts at Age Fifteen</i>	19
The frame of <i>The Mismeasure of Man</i> ,	19
Why revise <i>The Mismeasure of Man</i> after fifteen years?,	26
Reasons, history and revision of <i>The Mismeasure of Man</i> ,	36
1. Introduction	51
2. American Polygeny and Craniometry before Darwin: <i>Blacks and Indians as Separate, Inferior Species</i>	62
A shared context of culture,	63
Preevolutionary styles of scientific racism: monogenism and polygenism,	71
Louis Agassiz—America's theorist of polygeny,	74
Samuel George Morton—empiricist of polygeny,	82
<i>The case of Indian inferiority: Crania Americana</i>	
<i>The case of the Egyptian catacombs: Crania Aegyptiaca</i>	
<i>The case of the shifting black mean</i>	
<i>The final tabulation of 1849</i>	

Conclusions

The American school and slavery, 101

3. **Measuring Heads: *Paul Broca and the Heyday of Craniology*** **105**

The allure of numbers, 105

*Introduction**Francis Galton—apostle of quantification**A curtain-raiser with a moral: numbers do not guarantee truth*

Masters of craniometry: Paul Broca and his school, 114

*The great circle route**Selecting characters**Averting anomalies*

BIG-BRAINED GERMANS

SMALL-BRAINED MEN OF EMINENCE

LARGE-BRAINED CRIMINALS

FLAWS IN A PATTERN OF INCREASE THROUGH TIME

Front and back

THE CRANIAL INDEX

THE CASE OF THE FORAMEN MAGNUM

Women's brains

Postscript, 140

4. **Measuring Bodies: *Two Case Studies on the Apishness of Undesirables*** **142**

The ape in all of us: recapitulation, 142

The ape in some of us: criminal anthropology, 151

*Atavism and criminality**Animals and savages as born criminals**The stigmata: anatomical, physiological, and social**Lombroso's retreat**The influence of criminal anthropology**Coda*

Epilogue, 173

5. The Hereditarian Theory of IQ: An American Invention

176

Alfred Binet and the original purposes of the Binet scale, 176

Binet flirts with craniometry

Binet's scale and the birth of IQ

The dismantling of Binet's intentions in America

H. H. Goddard and the menace of the feeble-minded, 188

Intelligence as a Mendelian gene

GODDARD IDENTIFIES THE MORON

A UNILINEAR SCALE OF INTELLIGENCE

BREAKING THE SCALE INTO MENDELIAN COMPARTMENTS

THE PROPER CARE AND FEEDING (BUT NOT BREEDING) OF

MORONS

Preventing the immigration and propagation of morons

Goddard recants

Lewis M. Terman and the mass marketing of innate IQ, 204

Mass testing and the Stanford-Binet

Terman's technocracy of innateness

Fossil IQ's of past geniuses

Terman on group differences

Terman recants

R. M. Yerkes and the Army Mental Tests: IQ comes of age, 222

Psychology's great leap forward

Results of the army tests

A critique of the Army Mental Tests

THE CONTENT OF THE TESTS

INADEQUATE CONDITIONS

DUBIOUS AND PERVERSE PROCEEDINGS: A PERSONAL TESTIMONY

FINAGLING THE SUMMARY STATISTICS: THE PROBLEM OF ZERO VALUES

FINAGLING THE SUMMARY STATISTICS: GETTING AROUND OBVIOUS CORRELATIONS WITH ENVIRONMENT

Political impact of the army data

CAN DEMOCRACY SURVIVE AN AVERAGE MENTAL AGE OF THIRTEEN?

THE ARMY TESTS AND AGITATION TO RESTRICT IMMIGRATION:
 BRIGHAM'S MONOGRAPH ON AMERICAN INTELLIGENCE
 THE TRIUMPH OF RESTRICTION ON IMMIGRATION
 BRIGHAM RECANTS

6. The Real Error of Cyril Burt: *Factor Analysis and the Reification of Intelligence* **264**

- The case of Sir Cyril Burt, 264
 Correlation, cause, and factor analysis, 269
 Correlation and cause
 Correlation in more than two dimensions
 Factor analysis and its goals
 The error of reification
 Rotation and the nonnecessity of principal components
 Charles Spearman and general intelligence, 286
 The two-factor theory
 The method of tetrad differences
 Spearman's g and the great instauration of psychology
 Spearman's g and the theoretical justification of IQ
 Spearman's reification of g
 Spearman on the inheritance of g
 Cyril Burt and the hereditarian synthesis, 303
 The source of Burt's uncompromising hereditarianism
 BURT'S INITIAL "PROOF" OF INNATENESS
 LATER ARGUMENTS
 BURT'S BLINDNESS
 BURT'S POLITICAL USE OF INNATENESS
 Burt's extension of Spearman's theory
 Burt on the reification of factors
 Burt and the political uses of g
 L. L. Thurstone and the vectors of mind, 326
 Thurstone's critique and reconstruction
 The egalitarian interpretation of PMA's
 Spearman and Burt react

- Oblique axes and second-order g*
Thurstone on the uses of factor analysis
 Epilogue: Arthur Jensen and the resurrection of
 Spearman's *g*, 347
 A final thought, 350

7. A Positive Conclusion **351**

- Debunking as positive science, 351
 Learning by debunking, 352
 Biology and human nature, 354

Epilogue **365**

Critique of *The Bell Curve* **367**

- The Bell Curve*, 367
Disingenuousness of content
Disingenuousness of argument
Disingenuousness of program
 Ghosts of Bell Curves past, 379

Three Centuries' Perspectives on Race and Racism **391**

- Age-old fallacies of thinking and stinking, 391
 Racial geometry, 401
 The moral state of Tahiti—and of Darwin, 413

Bibliography **425**

Index **431**

The Real Error of Cyril Burt

Factor Analysis and the Reification of Intelligence

It has been the signal merit of the English school of psychology, from Sir Francis Galton onwards, that it has, by this very device of mathematical analysis, transformed the mental test from a discredited dodge of the charlatan into a recognized instrument of scientific precision.

—CYRIL BURT, 1921, p. 130

The case of Sir Cyril Burt

If I had any desire to lead a life of indolent ease, I would wish to be an identical twin, separated at birth from my brother and raised in a different social class. We could hire ourselves out to a host of social scientists and practically name our fee. For we would be exceedingly rare representatives of the only really adequate natural experiment for separating genetic from environmental effects in humans—genetically identical individuals raised in disparate environments.

Studies of identical twins raised apart should therefore hold pride of place in literature on the inheritance of IQ. And so it would be but for one problem—the extreme rarity of the animal itself. Few investigators have been able to rustle up more than twenty pairs of twins. Yet, amidst this paltriness, one study seemed to stand out: that of Sir Cyril Burt (1883–1971). Sir Cyril, doyen of mental testers, had pursued two sequential careers that gained him a preeminent role in directing both theory and practice in his field of educational psychology. For twenty years he was the official psychologist of the London County Council, responsible for the

administration and interpretation of mental tests in London's schools. He then succeeded Charles Spearman as professor in the most influential chair of psychology in Britain: University College, London (1932-1950). During his long retirement, Sir Cyril published several papers that buttressed the hereditarian claim by citing very high correlation between IQ scores of identical twins raised apart. Burt's study stood out among all others because he had found fifty-three pairs, more than twice the total of any previous attempt. It is scarcely surprising that Arthur Jensen used Sir Cyril's figures as the most important datum in his notorious article (1969) on supposedly inherited and ineradicable differences in intelligence between whites and blacks in America.

The story of Burt's undoing is now more than a twice-told tale. Princeton psychologist Leon Kamin first noted that, while Burt had increased his sample of twins from fewer than twenty to more than fifty in a series of publications, the average correlation between pairs for IQ remained unchanged to the third decimal place—a statistical situation so unlikely that it matches our vernacular definition of impossible. Then, in 1976, Oliver Gillie, medical correspondent of the London *Sunday Times*, elevated the charge from inexcusable carelessness to conscious fakery. Gillie discovered, among many other things, that Burt's two "collaborators," a Margaret Howard and a J. Conway, the women who supposedly collected and processed his data, either never existed at all, or at least could not have been in contact with Burt while he wrote the papers bearing their names. These charges led to further reassessments of Burt's "evidence" for his rigid hereditarian position. Indeed, other crucial studies were equally fraudulent, particularly his IQ correlations between close relatives (suspiciously too good to be true and apparently constructed from ideal statistical distributions, rather than measured in nature—Dorfman, 1978), and his data for declining levels of intelligence in Britain.

Burt's supporters tended at first to view the charges as a thinly veiled leftist plot to undo the hereditarian position by rhetoric. H. J. Eysenck wrote to Burt's sister: "I think the whole affair is just a determined effort on the part of some very left-wing environmentalists determined to play a political game with scientific facts. I am sure the future will uphold the honor and integrity of Sir Cyril without any question." Arthur Jensen, who had called Burt a

"born nobleman" and "one of the world's great psychologists," had to conclude that the data on identical twins could not be trusted, though he attributed their inaccuracy to carelessness alone.

I think that the splendid "official" biography of Burt recently published by L. S. Hearnshaw (1979) has resolved the issue so far as the data permit (Hearnshaw was commissioned to write his book by Burt's sister before any charges had been leveled). Hearnshaw, who began as an unqualified admirer of Burt and who tends to share his intellectual attitudes, eventually concluded that all allegations are true, and worse. And yet, Hearnshaw has convinced me that the very enormity and bizarreness of Burt's fakery forces us to view it not as the "rational" program of a devious person trying to salvage his hereditarian dogma when he knew the game was up (my original suspicion, I confess), but as the actions of a sick and tortured man. (All this, of course, does not touch the deeper issue of why such patently manufactured data went unchallenged for so long, and what this will to believe implies about the basis of our hereditarian presuppositions.)

Hearnshaw believes that Burt began his fabrications in the early 1940s, and that his earlier work was honest, though marred by rigid a priori conviction and often inexcusably sloppy and superficial, even by the standards of his own time. Burt's world began to collapse during the war, partly by his own doing to be sure. His research data perished in the blitz of London; his marriage failed; he was excluded from his own department when he refused to retire gracefully at the mandatory age and attempted to retain control; he was removed as editor of the journal he had founded, again after declining to cede control at the specified time he himself had set; his hereditarian dogma no longer matched the spirit of an age that had just witnessed the holocaust. In addition, Burt apparently suffered from Ménières disease, a disorder of the organs of balance, with frequent and negative consequences for personality as well.

Hearnshaw cites four instances of fraud in Burt's later career. Three I have already mentioned (fabrication of data on identical twins, kinship correlations in IQ, and declining levels of intelligence in Britain). The fourth is, in many ways, the most bizarre tale of all because Burt's claim was so absurd and his actions so patent and easy to uncover. It could not have been the act of a

rational man. Burt attempted to commit an act of intellectual parricide by declaring himself, rather than his predecessor and mentor Charles Spearman, as the father of a technique called "factor analysis" in psychology. Spearman had essentially invented the technique in a celebrated paper of 1904. Burt never challenged this priority—in fact he constantly affirmed it—while Spearman held the chair that Burt would later occupy at University College. Indeed, in his famous book on factor analysis (1940), Burt states that "Spearman's preeminence is acknowledged by every factorist" (1940, p. x).

Burt's first attempt to rewrite history occurred while Spearman was still alive, and it elicited a sharp rejoinder from the occupant emeritus of Burt's chair. Burt withdrew immediately and wrote a letter to Spearman that may be unmatched for deference and obsequiousness: "Surely you have a prior claim here. . . . I have been wondering where precisely I have gone astray. Would it be simplest for me to number my statements, then like my schoolmaster of old you can put a cross against the points where your pupil has blundered, and a tick where your view is correctly interpreted."

But when Spearman died, Burt launched a campaign that "became increasingly unrestrained, obsessive and extravagant" (Hearnshaw, 1979) throughout the rest of his life. Hearnshaw notes (1979, pp. 286–287): "The whisperings against Spearman that were just audible in the late 1930's swelled into a strident campaign of belittlement, which grew until Burt arrogated to himself the whole of Spearman's fame. Indeed, Burt seemed to be becoming increasingly obsessed with questions of priority, and increasingly touchy and egotistical." Burt's false story was simple enough: Karl Pearson had invented the technique of factor analysis (or something close enough to it) in 1901, three years before Spearman's paper. But Pearson had not applied it to psychological problems. Burt recognized its implications and brought the technique into studies of mental testing, making several crucial modifications and improvements along the way. The line, therefore, runs from Pearson to Burt. Spearman's 1904 paper was merely a diversion.

Burt told his story again and again. He even told it through one of his many aliases in a letter he wrote to his own journal and signed Jacques Lafitte, an unknown French psychologist. With the exception of Voltaire and Binet, M. Lafitte cited only English

sources and stated: "Surely the first formal and adequate statement was Karl Pearson's demonstration of the method of principal axes in 1901." Yet anyone could have exposed Burt's story as fiction after an hour's effort—for Burt never cited Pearson's paper in any of his work before 1947, while all his earlier studies of factor analysis grant credit to Spearman and clearly display the derivative character of Burt's methods.

Factor analysis must have been very important if Burt chose to center his quest for fame upon a rewrite of history that would make him its inventor. Yet, despite all the popular literature on IQ in the history of mental testing, virtually nothing has been written (outside professional circles) on the role, impact, and meaning of factor analysis. I suspect that the main reason for this neglect lies in the abstrusely mathematical nature of the technique. IQ, a linear scale first established as a rough, empirical measure, is easy to understand. Factor analysis, rooted in abstract statistical theory and based on the attempt to discover "underlying" structure in large matrices of data, is, to put it bluntly, a bitch. Yet this inattention to factor analysis is a serious omission for anyone who wishes to understand the history of mental testing in our century, and its continuing rationale today. For as Burt correctly noted (1914, p. 36), the history of mental testing contains two major and related strands: age-scale methods (Binet IQ testing), and correlational methods (factor analysis). Moreover, as Spearman continually stressed throughout his career, the theoretical justification for using a unilinear scale of IQ resides in factor analysis itself. Burt may have been perverse in his campaign, but he was right in his chosen tactic—a permanent and exalted niche in the pantheon of psychology lies reserved for the man who developed factor analysis.

I began my career in biology by using factor analysis to study the evolution of a group of fossil reptiles. I was taught the technique as though it had developed from first principles using pure logic. In fact, virtually all its procedures arose as justifications for particular theories of intelligence. Factor analysis, despite its status as pure deductive mathematics, was invented in a social context, and for definite reasons. And, though its mathematical basis is unassailable, its persistent use as a device for learning about the physical structure of intellect has been mired in deep conceptual errors from the start. The principal error, in fact, has involved a

major theme of this book: reification—in this case, the notion that such a nebulous, socially defined concept as intelligence might be identified as a “thing” with a locus in the brain and a definite degree of heritability—and that it might be measured as a single number, thus permitting a unilinear ranking of people according to the amount of it they possess. By identifying a mathematical factor axis with a concept of “general intelligence,” Spearman and Burt provided a theoretical justification for the unilinear scale that Binet had proposed as a rough empirical guide.

The intense debate about Cyril Burt’s work has focused exclusively on the fakery of his late career. This perspective has clouded Sir Cyril’s greater influence as the most powerful mental tester committed to a factor-analytic model of intelligence as a real and unitary “thing.” Burt’s commitment was rooted in the error of reification. Later fakery was the afterthought of a defeated man; his earlier, “honest” error has reverberated throughout our century and has affected millions of lives.

Correlation, cause, and factor analysis

Correlation and cause

The spirit of Plato dies hard. We have been unable to escape the philosophical tradition that what we can see and measure in the world is merely the superficial and imperfect representation of an underlying reality. Much of the fascination of statistics lies embedded in our gut feeling—and never trust a gut feeling—that abstract measures summarizing large tables of data must express something more real and fundamental than the data themselves. (Much professional training in statistics involves a conscious effort to counteract this gut feeling.) The technique of *correlation* has been particularly subject to such misuse because it seems to provide a path for inferences about causality (and indeed it does, sometimes—but only sometimes).

Correlation assesses the tendency of one measure to vary in concert with another. As a child grows, for example, both its arms and legs get longer; this joint tendency to change in the same direction is called a *positive correlation*. Not all parts of the body display such positive correlations during growth. Teeth, for example, do not grow after they erupt. The relationship between first incisor

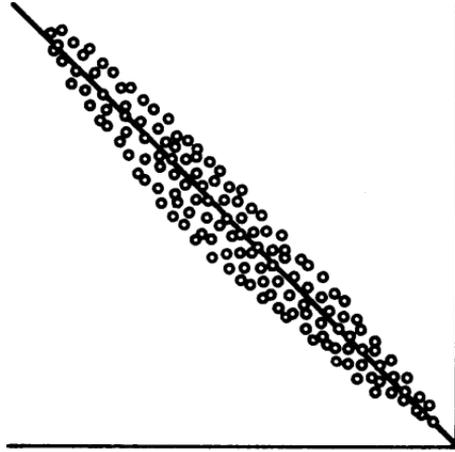
length and leg length from, say, age ten to adulthood would represent *zero correlation*—legs would get longer while teeth changed not at all. Other correlations can be negative—one measure increases while the other decreases. We begin to lose neurons at a distressingly early age, and they are not replaced. Thus, the relationship between leg length and number of neurons after mid-childhood represents *negative correlation*—leg length increases while number of neurons decreases. Notice that I have said nothing about causality. We do not know why these correlations exist or do not exist, only that they are present or not present.

The standard measure of correlation is called Pearson's product moment correlation coefficient or, for short, simply the correlation coefficient, symbolized as r . The correlation coefficient ranges from $+1$ for perfect positive correlation, to 0 for no correlation, to -1 for perfect negative correlation.*

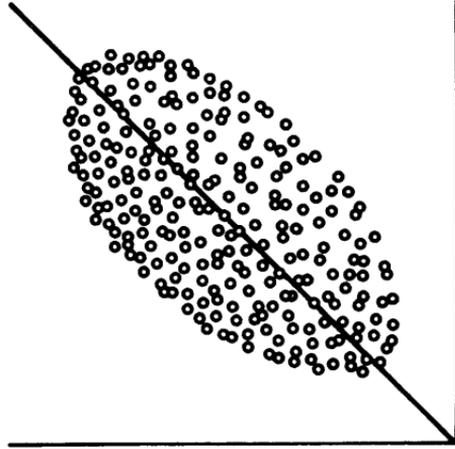
In rough terms, r measures the shape of an ellipse of plotted points (see Fig. 6.1). Very skinny ellipses represent high correlations—the skinniest of all, a straight line, reflects an r of 1.0 . Fat ellipses represent lower correlations, and the fattest of all, a circle, reflects zero correlation (increase in one measure permits no prediction about whether the other will increase, decrease, or remain the same).

The correlation coefficient, though easily calculated, has been plagued by errors of interpretation. These can be illustrated by example. Suppose that I plot arm length vs. leg length during the growth of a child. I will obtain a high correlation with two interesting implications. First, I have achieved *simplification*. I began with two dimensions (leg and arm length), which I have now, effectively, reduced to one. Since the correlation is so strong, we may say that the line itself (a single dimension) represents nearly all the information originally supplied as two dimensions. Secondly, I can, in this case, make a reasonable inference about the *cause* of this reduc-

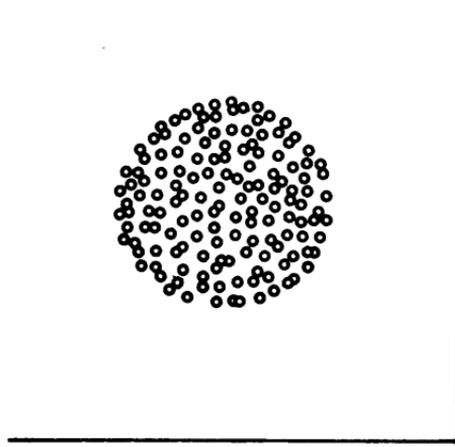
*Pearson's r is not an appropriate measure for all kinds of correlation, for it assesses only what statisticians call the intensity of linear relationship between two measures—the tendency for all points to fall on a single straight line. Other relationships of strict dependence will not achieve a value of 1.0 for r . If, for example, each increase of 2 units in one variable were matched by an increase in 2^2 units in the other variable, r would be less than 1.0 , even though the two variables might be perfectly "correlated" in the vernacular sense. Their plot would be a parabola, not a straight line, and Pearson's r measures the intensity of linear relationship.



high correlation



low correlation



no correlation

6 • 1 Strength of correlation as a function of the shape of an ellipse of points. The more elongate the ellipse, the higher the correlation.