IQ: THE RANK ORDERING OF THE WORLD

R. C. Lewontin, Steven Rose, and Leon J. Kamin

The Roots of IQ Testing

Social power runs in families. The probability that a child will grow into an adult in the highest 10 percent of income earners is ten times greater for children whose parents were in the top 10 percent than for children of the lowest 10 percent.[1] In France, the school failure rate of working-class children is four times that for children of the professional class.[2] How are we to explain hereditary differences in social power in a society that claims to have abolished hereditary privilege in the eighteenth century? One explanation—that hereditary privilege is integral to bourgeois society, which is not structurally conducive to real equality—is too disquieting and threatening; it breeds disorder and discontent; it leads to urban riots like those in Watts and Brixton. The alternative is to suppose that the successful possess an intrinsic merit, a merit that runs in the blood: Hereditary privilege becomes simply the ineluctable consequence of inherited ability. This is the explanation offered by the mental testing movement, whose basic argument can be summarized in a set of six propositions that, taken as a whole, form a seemingly logical explanation of social inequality. These are:

1. There are differences in status, wealth, and power.
2. These differences are consequences of different intrinsic ability, especially different "intelligence."
3. IQ tests are instruments that measure this intrinsic ability.
4. Differences in intelligence are largely the result of genetic differences between individuals.
5. Because they are the result of genetic differences, differences in ability are fixed and unchangeable.
6. Because most of the differences between individuals in ability are genetic, the differences between races and classes are also genetic and unchangeable.

While the argument begins with an undoubted truth that demands explanation, the rest is a mixture of factual errors and conceptual misunderstandings of elementary biology.

The purposes of Alfred Binet, who in 1905 published the first intelligence test, seem to have been entirely benign. The practical problem to which Binet addressed himself was to devise a brief testing procedure that could be used to help identify children who, as matters then stood, could not profit from instruction in the regular public schools of Paris. The problem with such children, Binet reasoned, was that their "intelligence" had failed to develop properly. The intelligence test was to be used as a diagnostic instrument. When the test had located a child with deficient intelligence, the next step was to increase the intelligence of such a child. That could be done, in Binet's view, with appropriate courses in "mental orthopedics." The important point is that Binet did not for a moment suggest that his test was a measure of some "fixed" or "innate" characteristic of the child. To those who asserted that the intelligence of an individual is a fixed quantity that one cannot augment, Binet's response was clear: "We must protest and react against this brutal pessimism."[3]

The basic principle of Binet's test was extraordinarily simple. With the assumption that the children to be tested had all shared a similar cultural background, Binet argued that older children should be able to perform mental tasks that younger children could not. To put matters very simply, we do not expect the average three-year-old to be able to recite the names of the months, but we do expect a normal ten-year-old to be able to do so. Thus, a ten-year-old who cannot recite the months is probably not very intelligent, while a three-year-old who can do so is probably highly intelligent. What Binet did, quite simply, was to put together sets of "intellectual" tasks appropriate for each age of childhood. There were, for example, some tasks that the average eight-year-old could pass, but which were too difficult for the average seven-year-old and very easy for the average nine-year-old. Those tasks defined the "mental age" of eight years. The intelligence of a child depended upon the relation his or her mental and chronological ages bore to each other. The child whose mental age was higher than his or her chronological age was "bright" or accelerated, and the child whose mental age was lower than his or her chronological age was "dull" or retarded. For most children, of course, the mental and chronological ages were the same. To Binet's satisfaction, the mental ages of children in a school class, as measured by his test, tended to correspond with teachers' judgments about which children were more or less "intelligent." That is scarcely surprising, since for the most part Binet's test involved materials and methods of approach similar to those emphasized in the school system. When a child lagged behind its age-mates by as much as two years of mental age, it seemed obvious to Binet that remedial intervention was called for. When two Belgian investigators reported that the children whom they had studied had much higher mental ages than the Paris children studied by Binet, Binet noted that the Belgian children attended a private school and came from the upper social classes. The small class sizes in the private school, plus the

kind of training given in a "cultured" home, could explain, in Binet's view, the higher intelligence of the Belgian children.

The translators and importers of Binet's test, both in the United States and in England, tended to share a common ideology, one dramatically at variance with Binet's. They asserted that the intelligence test measured an innate and unchangeable quantity, fixed by genetic inheritance. When Binet died prematurely in 1911, the Galtonian eugenicists took clear control of the mental testing movement in the English-speaking countries and carried their determinist principles even further. The differences in measured intelligence not just between individuals but between social classes and races were now asserted to be of genetic origin. The test was no longer regarded as a diagnostic instrument, helpful to educators, but could identify the genetically (and incurably) defective, those whose uncontrolled breeding posed a "menace . . . to the social, economic and moral welfare of the state."[4] When Lewis Terman introduced the Stanford-Binet test to the United States in 1916 he wrote that a low level of intelligence

> is very common among Spanish-Indian and Mexican families of the Southwest and also among negroes. Their dullness seems to be racial, or at least inherent in the family stocks from which they come. . . . The writer predicts that . . . there will be discovered enormously significant racial differences in general intelligence, differences which cannot be wiped out by any scheme of mental culture.
>
> Children of this group should be segregated in special classes. . . . They cannot master abstractions, but they can often be made efficient workers. . . . There is no possibility at present of convincing society that they should not be allowed to reproduce, although from a eugenic point of view they constitute a grave problem because of their unusually prolific breeding.[5]

Though Terman's Stanford-Binet test was basically a translation of Binet's French items, it contained two significant modifications. First, a set of items said to measure the intelligence of adults was included, as well as items for children of different ages. Second, the ratio between mental and chronological age, the "intelligence quotient," or IQ, was now calculated to replace the simple statement of mental and chronological ages. The clear implication was that the IQ, fixed by the genes, remained constant throughout the individual's life. "The fixed character of mental levels" was cited by another translator of Binet's test, Henry Goddard, in a 1919 lecture at Princeton University, as the reason why some were rich and others poor, some employed and others unemployed. "How can there be such a thing as social equality with this wide range of mental capacity? . . . As for an equal distribution of the wealth of the world, that is equally absurd."[6]

The major translator of Binet's test in England was Cyril Burt, whose links to Galtonian eugenics were even more pronounced than those of his American contemporaries. Burt's father was a physician who treated Galton, and Galton's strong recommendations hastened Burt's appointment as the first school psychologist in the English-speaking world. As early as 1909 Burt had administered some crude tests to two very small groups of schoolchildren in the town of Oxford. The children at one school were the sons of Oxford dons, fellows of the Royal Society, etc., while children at the other school were the sons of ordinary townspeople. Burt claimed that the children from the higher-class school did better on his tests and that this demonstrated that intelligence was inherited. This scientifically stated conclusion, published in the 1909 *British Journal of Psychology,*[7] might have been predicted from Burt's handwritten entry, six years earlier, in his Oxford undergraduate notebook: "The problem of the very poor—chronic poverty: Little prospect of the solution of the problem without the forcible detention of the wreckage of society or other preventing them from propagating their species."

Burt continued his eugenic researches into the inheritance of IQ until he died in 1971, knighted by his monarch and bemedaled by the American Psychological Association. The masses of data that he published helped to establish the "eleven-plus" examination in England, linked to the postwar system of selective education. "Intelligence," Burt wrote in 1947, "will enter into everything the child says, thinks, does or attempts, both while he is at school and later on. . . . If intelligence is innate, the child's degree of intelligence is permanently limited." Further, "Capacity must obviously limit content. It is impossible for a pint jug to hold more than a pint of milk; and it is equally impossible for a child's educational attainments to rise higher than his educable capacity permits."[8] There could be no clearer statement of what had happened to Binet's test in the hands of the Galtonians. The test designed to alert educators that they must intervene with special educational treatment was now said to measure "educable capacity." When a child did poorly in school, or when an adult was unemployed, it was because he or she was genetically inferior and must always remain so. The fault was not in the school or in the society, but in the inferior person.

The IQ test, in practice, has been used both in the United States and England to shunt vast numbers of working-class and minority children onto inferior and dead-end educational tracks.* The reactionary impact of the test, however, has extended far beyond the classroom. The testing movement was clearly linked, in the United States, to the passage, beginning in 1907, of compulsory sterilization laws aimed at genetically inferior "degenerates." The categories detailed included, in different states, criminals, idiots, imbeciles, epileptics, rapists, lunatics, drunkards, drug fiends, syphilitics, moral and sexual perverts, and "diseased and degenerate persons." The sterilization laws, explicitly declared constitutional by the U.S. Supreme Court in 1927, established as a matter of legal fact the core assertion of biological determinism: that all these degenerate characteristics were transmitted through the genes. When the IQ testing program of the United States Army in World War I indicated that immigrants from Southern and Eastern Europe had low test scores, this was said to demonstrate that "Alpines" and "Mediterraneans" were genetically inferior to "Nordics." The army of IQ data figured prominently in the public and congressional debates over the Immigration Act of 1924. That overtly racist act established as a feature of American

*"Tracking" in the U.S. educational system is more or less synonymous with "streaming" in Britain.

immigration policy a system of "national origin quotas." The purpose of the quotas was explicitly to debar, as much as possible, the genetically inferior peoples of Southern and Eastern Europe, while encouraging "Nordic" immigration from northern and western Europe. This tale has been told in full elsewhere.[9]

Today many (if not most) psychologists recognize that differences in IQ between various races and/or ethnic groups cannot be interpreted as having a genetic basis. The obvious fact is that human races and populations differ in their cultural environments and experiences, no less than in their gene pools. There is thus no reason to attribute average score differences between groups to genetic factors, particularly since it is so obviously the case that the ability to answer the kinds of questions asked by IQ testers depends heavily on one's past experience. Thus, during World War I, the Army Alpha test asked Polish, Italian, and Jewish immigrants to identify the product manufactured by Smith & Wesson and to give the nicknames of professional baseball teams. For immigrants who could not speak English, the Army Beta test was designed as a "nonverbal" measure of "innate intelligence." That test asked the immigrants to point out what was missing from each of a set of drawings. The set included a drawing of a tennis court, with the net missing. The immigrant who could not answer such a question was thereby shown to be genetically inferior to the tennis-playing psychologists who devised such tests for adults.

What IQ Tests Measure

How do we know that IQ tests measure "intelligence"? Somehow, when the tests are created, there must exist a prior criterion of intelligence against which the results of the tests can be compared. People who are generally considered "intelligent" must rate high and those who are obviously "stupid" must do badly or the test will be rejected. Binet's original test, and its adaptations into English, were constructed to correspond to teachers' and psychologists' a priori notions of intelligence. Especially in the hands of Terman and Burt, they were tinkered with and standardized so that they became consistent predictors of school performance. Test items that differentiated boys from girls, for example, were removed, since the tests were not meant to make that distinction; differences between social classes, or between ethnic groups or races, however, have not been massaged away, precisely because it is these differences that the tests are *meant* to measure.

IQ tests at present vary considerably in their form and content, but all of them are validated by how well they agree with older standards. It must be remembered that an IQ test is published and distributed by a publishing company as a commercial item, selling hundreds of thousands of copies. The chief selling point of such tests, as announced in their advertising, is their excellent agreement with the results of the Stanford-Binet test. Most combine tests of vocabulary, numerical reasoning, analogical reasoning, and pattern recognition.

Some are filled with specific and overt cultural references: Children are asked to identify characters from English literature ("Who was Wilkins Micawber?"); they are asked to make class judgments ("Which of the five persons below is most like a carpenter, plumber, and bricklayer? 1) postman, 2) lawyer, 3) truck driver, 4) doctor, 5) painter"); they are asked to judge socially acceptable behavior ("What should you do when you notice you will be late to school?"); they are asked to judge social stereotypes ("Which is prettier?" when given the choice between a girl with some Negroid features and a doll-like European face); they are asked to define obscure words (sudorific, homunculus, parterre). Of course, the "right" answers to such questions are good predictors of school performance.

Other tests are "nonverbal" and consist of picture explanations or geometric pattern recognition. All—and most especially the nonverbal tests—depend upon the tested person having learned the ability to spend long periods participating in a contentless, contextless mental exercise under the supervision of authority and under the implied threat of reward or punishment that accompanies all tests of any nature. Again, they necessarily predict school performance, since they mimic the content and circumstances of schoolwork.

IQ tests, then, have not been designed from the principles of some general theory of intelligence and subsequently shown to be independently a predictor of social success. On the contrary, they have been empirically adjusted and standardized to correlate well with school performance, while the notion that they measure "intelligence" is added on with no independent justification to validate them. Indeed, we do not know what that mysterious quality "intelligence" is. At least one psychologist, E. G. Boring, has defined it as "what intelligence tests measure."[10] The empirical fact is that there exist tests that predict reasonably well how children will perform in school. That these tests advertise themselves as "intelligence" measures should not delude us into investing them with more meaning than they have.

Reifying Behavior

The possibility of behavioral measurements rests upon certain basic underlying assumptions, which should now be clarified. First, it is assumed that it is possible to define, absolutely or operationally, a particular "quality" to be measured. Some such qualities, like height, are relatively unproblematic. To the question, "How tall are you?" the answer in centimeters, feet, or inches is easy to give. To the question "How angry are you?" no such easy answer can be given. Anger has to be defined operationally, as, for instance, how often an individual placed in a given test situation and asked the question by the experimenter responds by hitting him on the nose. This is not a flippant example. "Aggression" in a rat is measured by putting a mouse in a cage with it and observing the behavior and time taken for the rat to kill the mouse. Sometimes this is described under the name "muricidal" behavior in the literature, which presumably makes the exper-

imenters happier that they are measuring something really scientific. Research in this area thus becomes forced into Boring's circularity: Intelligence "is" what intelligence tests measure.

The "quality" is then taken to be an underlying object that is merely reflected in varying aspects of an individual's behavior under widely different circumstances. Thus "aggression" is what individuals express when a man beats his wife, pickets boycott scabs during a strike, teenagers fight after a football game, black Africans struggle for independence from their colonial masters, generals press buttons unleashing thermonuclear war, or America and the Soviet Union compete in the Olympic Games or the space race. The underlying quality is identical with that which underlies muricide in rats.

Second, it is assumed that the quality is a fixed property of an individual. Aggression and intelligence are seen not as processes that emerge from a situation and are part of the relationships of that situation, but rather exist like reservoirs each of defined amount, inside each of us, to be turned on or off. Instead of seeing the anger or aggression expressed in inner city riots as emerging from the interaction between individuals and their social and economic circumstances and as expressive of collective action—therefore a social phenomenon—the biological determinist argument defines inner city violence as merely the sum of individual units of aggressiveness. . . .

Thus verbs are redefined as nouns: processes of interaction are reified and located inside the individual. Further, reified verbs, like aggression, are assumed to be rigid, fixed things that can be reproducibly measured. Like height, they will not vary much from day to day; indeed, if the tests designed to measure them show such variations they are regarded as poor tests. It is assumed not that the "quality" being measured is labile, but that our instruments need greater precision.

Psychometry and the Obsession with the Norm

Implicit in reification is the third and crucial premise of the mental testing movement. If processes are really things that are the properties of individuals and that can be measured by invariant objective rules, then there must be scales on which they can be located. The scale must be metric in some manner, and it must be possible to compare individuals across the scale. If one person has an aggression score of 100 and the next of 120, the second is therefore 20 percent more aggressive than the first. The fault in the logic should be clear: The fact that it is possible to devise tests on which individuals score arbitrary points does not mean that the quality being measured by the test is really metric. The illusion is provided by the scale. Height is metric, but consider, for instance, color. We could present individuals with a set of colors ranging from red to blue and ask them to rank them as 1 (reddest) to 10 (bluest). But this would not mean that the color rated 2 was actually twice as blue as the color rated 1. The ordinal scale is an arbitrary one, and most psychometric tests are actually ordinals of this sort. If one rat kills ten mice in five minutes and a second rat kills twelve in the same time, this does not automatically mean that the second is 20 percent more aggressive than the first. If one student scores 80 in an exam and a second 40, this does not mean the first is twice as intelligent as the second.

Surmounting or disguising the scaling problem is integral to the grand illusion of psychometry. Individuals vary in height, and if heights for a hundred or so individuals drawn at random from a population are plotted, they will likely fall into the normal distribution, or bell-shaped curve. If the divisions in one's scale are very fine—say, inches—the bell curve is quite wide. If we had no measures less than feet, and we measured each individual to the nearest foot, the curve would be much narrower at the bottom. The vast majority of individuals in Western society would lie between the five- and six-foot measure. While we know the relationship of inches to feet and could under appropriate circumstances convert from one scale to another, and we know when to use each, as when we are finding a pair of shoes that fit or deciding the best size to make a door opening, we do not know the comparable relationships between different ways of measuring aggression or intelligence. Which scale is chosen depends on whether one wants to make differences of scale appear large or small, and these decisions are those that psychometry arbitrarily makes. The decision that a "good" scale is one in which two-thirds of the population should lie within 15 percent of the mean score of the entire population—the famous normal distribution—is arbitrary, but its power is such that psychometrists chop and change their scales till they meet this criterion.

Yet the power of the "norm," once established, is that it is used to judge individuals who have been located along its linear scale. Deviations from the norm are regarded with alarm. Parents who are told that their child is two standard deviations from the norm on some behavioral score are led to believe that he or she is "abnormal" and should be adjusted in some way to psychometry's Procrustean bed. Psychometry, above all, is a tool of a conformist society that, for all its professed concern with individuals, is in reality mainly concerned to match them against others and to attempt to adjust them to conformity.

Pressure to conform to social norms, and institutions that propagate and reinforce these norms, are, of course, characteristic of all human societies. In advanced capitalist societies and today's state capitalist societies like the Soviet Union or those in Eastern Europe, the norm becomes an ideological weapon in its own right, foreshadowed by Huxley's *Brave New World* and Orwell's *1984* but cloaked in the benign language of those who only wish to help, to advise, but not to control and manipulate. Let us be clear: norms are statistical artifacts; they are not biological realities. Biology is not committed to bell-shaped curves.

IQ Tests as Predictors of Social Success

The claim that IQ tests are good predictors of eventual social success is, except in a trivial and misleading sense, simply incorrect. It is true that if one measures

social success by income or by what sociologists call socioeconomic status (SES)—a combination of income, years of schooling, and occupation—then people with higher incomes or higher SES did better on IQ tests when they were children than did people with low incomes or low SES. For example, a person whose childhood IQ was in the top 10 percent of all children is fifty times more likely to wind up in the top 10 percent of income than a child whose IQ was in the lowest 10 percent. But that is not really quite the question of interest. What we really should ask is: How much more likely is a high-IQ child to wind up in the top 10 percent of income, *all other things being equal?* In other words, there are multiple and complex causes of events which do not act or exist independently of each other. Even where *A* looks at first sight as if it is the cause of *B*, it sometimes really turns out on deeper examination that *A* and *B* are both effects of some prior cause, *C*. For example, on a worldwide basis, there is a strong positive relationship between how much fat and how much protein the population of a particular country consumes. Rich countries consume a lot of each, poor countries little. But fat consumption is neither the cause nor the result of eating protein. Both are the consequence of how much money people have to spend on food. Thus, although fat consumption per capita is statistically a predictor of protein consumption per capita, it is not a predictor when all other things are equal. Countries that have the same per capita income show no particular relation between average fat and average protein consumption, since the real causal variable, income, is not varying between countries.

This is precisely the situation for IQ performance and eventual social success. They go together because both are the consequences of other causes. To see this, we can ask how good a predictor IQ is of eventual economic success when we hold constant the person's family background and the number of years of schooling. With these constant, a child in the top 10 percent of IQ has only twice, not fifty times, the chance of winding up in the top 10 percent of income as a child of the lowest IQ group. Conversely, and more important, a child whose family is in the top 10 percent of economic success has a 25 times greater chance of also being at the top than the child of the poorest 10 percent of families, even when both children have average IQ.[11] Family background, rather than IQ, is the overwhelming reason why an individual ends up with a higher than average income. Strong performance on IQ tests is simply a reflection of a certain kind of family environment, and once that latter variable is held constant, IQ becomes only a weak predictor of economic success. If there is indeed an intrinsic ability that leads to success, IQ tests do not measure it. If IQ tests do measure intrinsic intelligence as is claimed, than clearly it is better to be born rich than smart.

The Heritability of IQ

The next step in the determinist argument is to claim that differences between individuals in their IQ arise from differences in their genes. The notion that intelligence is hereditary is, of course, deeply built into the theory of IQ testing itself because of its commitment to the measurement of something that is intrinsic and unchangeable. From the very beginning of the American and British mental testing movement, it was assumed that IQ was biologically heritable.

There are certain erroneous senses of "heritable" that appear in the psychometricians' writings on IQ, mixed up with the geneticists' technical meaning of heritability, and which contribute to false conclusions about the consequences of heritability. The first error is that genes themselves determine intelligence. Neither for IQ nor for any other trait can genes be said to determine the organism. There is no one-to-one correspondence between the genes inherited from one's parents and one's height, weight, metabolic rate, sickness, health, or any other nontrivial organic characteristic. The critical distinction in biology is between the *phenotype* of an organism, which may be taken to mean the total of its morphological, physiological, and behavioral properties, and its *genotype*, the state of its genes. It is the genotype, not the phenotype, that is inherited. The genotype is fixed; the phenotype develops and changes constantly. The organism itself is at every stage the consequence of a developmental process that occurs in some historical sequence of environments. At every instant in development (and development goes on until death) the next step is a consequence of the organism's present biological state, which includes both its genes and the physical and social environment in which it finds itself. This comprises the first principle of developmental genetics: that every organism is the unique product of the interaction between genes and environment at every stage of life. While this is a textbook principle of biology, it has been widely ignored in determinist writings. "In the actual race of life, which is not to get ahead, but to get ahead of somebody," wrote E. L. Thorndike, the leading psychologist of the first half of the century, "the chief determining factor is heredity."[12]

The second error—even if admitting that genes do not determine the actual developmental outcome—is to claim that they determine the effective limit to which it can go. Burt's metaphor of the pint jug that can hold no more than a pint of milk is a precise image of this view of genes as the determinants of capacity. If the genetic capacity is large, the argument runs, then an enriched environment will result in a superior organism, although in a poor environment the same individual will not show much ability. If the genetic capacity is poor, however, then an enriched environment will be wasted. Like the notion of the absolute determination of organisms by genes, this view of genetic "capacity" is simply false. There is nothing in our knowledge of the action of genes that suggests differential total capacity. In theory, of course, there must be *some* maximum height, say, to which an individual could grow; but in fact there is no relationship between that purely theoretical maximum, which is never reached in practice, and the actual variations among individuals. The lack of relationship between actual state and theoretical maximum is a consequence of the fact that growth rates and growth maxima are not related. Sometimes it is the slowest growers that reach the greatest size. The proper description of the difference between genetic types is not in some hypothetical "capacity" but in the specific pheno-

type that will develop for that genotype as a consequence of some specific chain of environmental circumstances.

Nor, of course does the phenotype develop linearly from the genotype from birth to adulthood. The "intelligence" of an infant is not merely a certain small percentage of that of the adult it will become, as if the "pint jug" were being steadily filled. The process of growing up is not a linear progression from incompetence to competence: To survive, a newborn baby must be competent at being a newborn baby, not at being a tiny version of the adult it will later become. Development is not just a quantitative process but one in which there are transformations in quality—between suckling and chewing solid food, for instance, or between sensorimotor and cognitive behavior. But such transitions are not permitted in the rank-ordered view of the universe that determinism offers.

The total variation in phenotype in a population of individuals arises from two interacting sources. First, individuals with the same genes still differ from each other in phenotype because they have experienced different developmental environments. Second, there are different genotypes in the population which differ from each other on the average even in the same array of environments. The phenotype of an individual cannot be broken down into the separate contributions of genotype and of environment, because the two interact to produce the organism; but the total variation of any phenotype in the population can be broken down into the variation between the average of the different genotypes and the variation among individuals with the same genotype. The variation between the average performance of different genotypes is called the *genetic variance* of the trait (that is, the aspect of the phenotype under study—eye color, height, or whatever) in the population, while the variation among individuals of the same genotype is called the *environmental variance* of the trait in the population. It is important to notice that the genetic and environmental variances are not universal properties of a trait but depend upon which population of individuals is being characterized and under which set of environments. Some populations may have a lot of genetic variance for a character, some only a little. Some environments are more variable than others.

The *heritability* of a trait, in the technical sense in which geneticists understand it, is the proportion of all the variation of a trait in a population that is accounted for by the genetic variance. Symbolically,

$$\text{Heritability} = \text{H} = \frac{\text{genetic variance}}{\text{genetic variance} + \text{environmental variance}}$$

If the heritability is 100 percent, then all of the variance in the population is genetic. Each genotype would be phenotypically different, but there would be no developmental variation among individuals of the same genotype. If the heritability is zero, all of the variation is among individuals within a genotype, and there is no average variation from genotype to genotype. Characters like height, weight, shape, metabolic activity, and behavioral traits all have heritabilities below 100 percent. Some, like specific language spoken or religious and political affiliation, have heritabilities of zero. The claim of biological determinists has been that the heritability of IQ is about 80 percent. How do they arrive at this figure?

Estimating the Heritability of IQ

All genetic studies are studies of the resemblances of relatives. If a trait is heritable, that is, if different genotypes have different average performances, then relatives ought to resemble each other more closely than unrelated persons do, since relatives share genes from common ancestors. Brothers and sisters ought to be more like each other than aunts and nephews, who ought to be more similar than totally unrelated people. The standard measure of similarity between things that vary quantitatively is their *correlation*, which measures the degree to which larger values for one variable go together with larger values of a second variable, and smaller values with smaller values. The correlation coefficient, r, ranges from $+1.0$ for perfect positive correlation, through zero for no relationship, to -1.0 for perfect negative correlation. So, for example, there is a positive correlation between father's income and child's years of schooling. Richer fathers have better-educated children while poorer fathers have less-educated children, on the average. The correlation is not perfect, since some families produce children who go to graduate school, but it is positive. In contrast, in the United States there is a negative correlation between family income and the number of visits per year to hospital emergency rooms. The lower your income, the more likely you are to use the emergency room as a medical service instead of a private doctor.

One important point about correlation is that it measures how two things vary together but does not measure how similar their average levels are. So the correlation between the heights of mothers and their sons could be perfect in that taller mothers had the taller sons and shorter mothers had the shorter sons, yet all the sons could be taller than all the mothers. Covariation is not the same as identity. The significance of this fact for the heritability of IQ and its meaning is considerable. Suppose a group of fathers had IQs of 96, 97, 98, 99, 100, 101, 102, and 103, while their daughters, separated from their fathers at birth and raised by foster parents, had IQs respectively of 106, 107, 108, 109, 110, 111, 112, and 113. There is a perfect correspondence between the IQs of fathers and daughters, and we might judge the character to be perfectly heritable because, knowing a father's IQ, we could tell without error which of the daughters was his. The correlation is, in fact, $r = +1.0$, yet the daughters are ten points above their fathers in IQ, so the experience of being raised by foster parents had a powerful effect. There is thus no contradiction between the assertion that a trait is perfectly heritable and the assertion that it can be changed radically by environment. As we shall see, this is not a hypothetical example.

Second, a correlation between two variables is not a reliable guide to causation. If *A* and *B* are correlated, one may be the cause of the other, they may both

be the consequence of a common cause, or they may be entirely accidentally related. The number of cigarettes smoked per day is correlated with the chance of lung cancer because smoking is a cause of lung cancer. The floor area of a person's house and the average age to which he or she lives are positively correlated not because living in a big house is conducive to health but because both characteristics are a consequence of the same cause—high income. For that matter, the distance of the Earth from Halley's comet and the price of fuel are negatively correlated in recent years because one has been decreasing while the other increased, but for totally independent reasons.

In general, heritability is estimated from the correlation of a trait between relatives. Unfortunately, in human populations two important sources of correlation are conflated: Relatives resemble each other not only because they share genes but also because they share environments. This is a problem that can be circumvented in experimental organisms, where genetically related individuals can be raised in controlled environments, but human families are not rat cages. Parents and their offspring may be more similar than unrelated persons because they share genes but also because they share family environment, social class, education, language, etc. To solve this problem, human geneticists and psychologists have taken advantage of special circumstances that are meant to break the tie between genetic and environmental similarity in families.

The first circumstance is adoption. Are particular traits in adopted children correlated with their biological families even when they have been separated from them? Are identical (i.e. monozygotic, or one-egg) twins, separated at birth, similar to each other in some trait? If so, genetic influence is implicated. The second circumstance holds environment constant but changes genetic relationship. Are identical twins more alike than fraternal (i.e. dizygotic, or two-egg) twins? Are two biological brothers or sisters (sibs) in a family more alike than two adopted children in a family? If so, genes are again implicated because, in theory, identical twins and fraternal twins have equal environmental similarity but they are not equally related genetically.

The difficulty with both these kinds of observations is that they only work if the underlying assumptions about environment are true. For the adoption studies to work, it must be true that there is no correlation between the adopting families and the biological families. There must not be selective placement of adoptees. In the case of one-egg and two-egg twins, it must be true that identical twins do not experience a more similar environment than fraternal twins. As we shall see, these problems have been largely ignored in the rush to demonstrate the heritability of IQ.

The theory of estimating heritability is very well worked out. It is well known how large samples should be to get reliable estimates. The designs of the observations to avoid selective adoptions, to get objective measures of test performance without bias on the part of the investigator, to avoid statistical artifacts that may arise from unrepresentative samples of adopting families, are all well laid out in textbooks of statistics and quantitative genetics. Indeed, these theories are constantly put into practice by animal breeders who would be unable to have

their research reports published in genetics journals unless they adhered strictly to the standard methodological requirements. The record of psychometric observations on the heritability of IQ is in remarkable contrast. Inadequate sample sizes, biased subjective judgments, selective adoption, failure to separate so-called "separated twins," unrepresentative samples of adoptees, and gratuitous and untested assumptions about similarity of environments are all standard characteristics in the literature of IQ genetics. There has even been, as we shall see, massive and influential fraud. We will review in some detail the state of psychometric genetic observations—not simply because it calls into question the actual heritability of IQ, but because it raises the far more important issue of why the canons of scientific demonstration and credibility should be so radically different in human genetics than in the genetics of pigs. Nothing demonstrates more clearly how scientific methodology and conclusions are shaped to fit ideological ends than the sorry state of the heritability of IQ.

The Cyril Burt Scandal

The clearest evidence, by far, for the genetic determination of IQ was the massive life's work of the late Sir Cyril Burt. In 1969 Arthur Jensen quite correctly referred to Burt's work as "the most satisfactory attempt" to estimate the heritability of IQ. When Burt died, Jensen referred to him as "a born nobleman," whose "larger, more representative samples than any other investigator in the field has ever assembled" would secure his "place in the history of science."[13] Hans Eysenck wrote that he drew "rather heavily" on Burt's work, citing "the outstanding quality of the design and the statistical treatment in his studies."[14]

The Burt data seemed so impressive for a number of very good reasons. First, one of the simplest ways, at least in theory, of demonstrating the heritable basis of a trait is to study separated identical twins. The separated twin pairs have identical genes, and they are assumed not to have shared any common environment. Thus, if they resemble one another markedly in some respect, the resemblance must be due to the only thing they share in common: their identical genes. The largest IQ study of separated identical twins ever reported, supposedly based on fifty-three twin pairs, was that of Cyril Burt. The IQ correlation of separated twin pairs reported by Burt was strikingly high, more so than that reported in the three other studies of separated twins. The most important aspect of Burt's study, however, was that he alone had been able to measure quantitatively the similarity of the environments in which the separated twin pairs had been reared. The incredible (and convenient) result reported by Burt was that there was no correlation at all between the environments of the separated pairs.

Further, in order to fit a genetic model to IQ data, it is necessary to know what the IQ correlations are for a considerable number of types of relatives—some close and some not so close. Burt was the only investigator in history who claimed to have administered the same IQ test, in the same population, to the full

gamut of biological relatives of all degrees of closeness. In fact, for some types of relatives (grandparent-grandchild, uncle-nephew, second cousin pairs), the IQ correlations reported by Burt are the *only* such correlations *ever* to have been reported. The Burt correlations for all types of relatives corresponded with remarkable precision to the values expected if IQ were almost entirely determined by the genes.

The blunt fact is that Burt's data, which have played so important a role, were reported and published in what is clearly a truly scandalous and suspicious fashion. The implausibility of Burt's claims should have been noted at once by any reasonably alert and conscientious scientific reader. To begin with, Burt never provided even the most elementary description of how, when, or where his "data" had been collected. The normal canons of scientific reporting were ignored entirely by Burt, and by the editors of the journals that published his papers. He never even identified the "IQ test" he supposedly administered to untold thousands of pairs of relatives. Within many of his papers, even the sizes of his supposed samples of relatives were not reported. The correlations were given without any supporting details. The 1943 paper that first reported many of the correlations between relatives made only the following reference to procedural details: "Some of the inquiries have been published in LCC reports or elsewhere; but the majority remain buried in typed memoranda or degree theses."[15] Conscientious scientists usually do not refer interested readers to their primary sources and documentation in such a cavalier way. The reader should not be surprised by the fact that none of the London County Council reports, typed memoranda, or degree theses glancingly referred to by Burt have ever come to light.

The very few occasions when Burt made specific statements about his procedure should have provoked some doubts in his scientific readers. For example, in a 1955 paper Burt described the procedure by which he obtained IQ test results for parent-child, grandparent-grandchild, uncle-nephew, etc. The IQ data for children were supposedly obtained by revising (on the basis of teacher's comments) the results of unspecified IQ tests given in school. But how did Burt obtain "IQs" for adults? He wrote: "For the assessments of the parents we relied chiefly on personal interviews; but in doubtful or borderline cases an open or a camouflaged test was employed."[16] That is, in measuring the "IQs" of adults Burt did not even *claim* to have administered an objective, standardized IQ test. The IQ was said to have been guessed at during an interview! The spectacle of Professor Burt administering "camouflaged" IQ tests while chatting with London grandparents is the stuff of farce, not of science. The correlations reported by Burt on this claimed basis, however, were routinely presented as hard scientific truths in textbooks of psychology, of genetics, and of education. Professor Jensen referred to precisely this work as "the most satisfactory attempt" to estimate the heritability of IQ. When Burt's procedure was publicly criticized, Hans Eysenck was able to write in Burt's defense: "I could only wish that modern workers would follow his example."[17]

The collapse of Burt's claims within the scientific community began when

attention was drawn to some numerical impossibilities in Burt's published papers.[18] For example, Burt in 1955 claimed to have studied twenty-one pairs of separated identical twins and reported that, on some unnamed group test of intelligence, their IQ correlation was .771. By 1958 the number of pairs had been increased to "over 30"; surprisingly, the IQ correlation remained precisely .771. By 1966, when the sample size had been increased to fifty-three pairs, the correlation was still exactly .771! This remarkable tendency for IQ correlations to remain identical to the third decimal place was also true of Burt's studies of nonseparated identical twin pairs: as the sample size increased progressively with time, the correlation failed to change. The same identity to the third decimal place was also true of IQ correlations for other types of relatives published by Burt, as sample sizes increased (or in some cases decreased) over time. These and other characteristics indicated that, at the very least, Burt's data and claimed results could not be taken seriously. As one of us in 1974 concluded after surveying Burt's work: "The numbers left behind by Professor Burt are simply not worthy of our current scientific attention."[19]

The scientific exposure of Burt prompted Professor Jensen to execute a brisk about-face. Two years earlier Jensen had described Burt as a born nobleman, whose large and representative samples had secured his place in the history of science. But in 1974 Jensen wrote, after citing the absurdities that critics had already documented, that Burt's correlations were "useless for hypothesis testing"—that is to say, worthless.[20] But Jensen maintained that Burt's work had merely been careless, not fraudulent; and he also maintained that the elimination of Burt's data did not substantially reduce the weight of the evidence demonstrating a high heritability of IQ. That incredible claim was made despite Jensen's earlier assertion that Burt's was "the most satisfactory attempt" to calculate the heritability of IQ.[21]

The argument over Burt's data might have remained a discreet academic affair and might have tiptoed around the question of Burt's fraudulence were it not for the medical correspondent of the London *Sunday Times*, Oliver Gillie. Gillie tried to locate two of Burt's research associates, the Misses Conway and Howard, who had supposedly published papers in a psychological journal edited by Burt. According to Burt, they were responsible for the IQ testing of the separated identical twins, for the testing of other types of relatives, and for much of Burt's published data analyses. But Gillie could uncover absolutely no documentary record of the existence of these research associates. They had not been seen by, and were wholly unknown to, Burt's closest co-workers. When asked about them by his housekeeper, Burt had replied that they had emigrated to Australia or New Zealand, this at a time *before*, according to Burt's published papers, they were testing twins in England. Burt's secretary indicated that Burt had sometimes written papers signed by either Conway or Howard. These facts led Gillie to suggest, in a front-page article in 1976, that Conway and Howard may never have existed.[22] The article flatly accused Burt of perpetrating a major scientific fraud, a charge subsequently supported by two of Burt's former students, now themselves prominent psychometricians, Alan and Ann Clarke.

The public exposure of Burt's fraudulence seemed to strike a raw hereditarian nerve. Professor Jensen wrote that the attack on Burt was designed "to wholly discredit the large body of research on the genetics of human mental abilities. The desperate scorched-earth style of criticism 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 . . . take warranted legal action against such unfounded defamation."[23] Professor Eysenck joined in by pointing out that Burt had been "knighted for his services" and that the charges against him contained "a whiff of McCarthyism, of notorious smear campaigns, and of what used to be known as character assassination."[24]

The attempt to defend Burt by assaulting his critics soon collapsed. The eulogy at Burt's memorial service had been delivered by an admirer, Professor Leslie Hearnshaw, and had prompted Burt's sister, in 1971, to commission Hearnshaw to write a biography of her distinguished brother and to make Burt's private papers and diaries freely available to him. When the fraud charges exploded, Hearnshaw wrote to the *Bulletin* of the British Psychological Society, indicating that he would assess all the available evidence and warning that the charges of Burt's critics could not be lightly dismissed. This warning seems to have muted the tone of Burt's more militant hereditarian defenders. Thus, by 1978, Eysenck wrote of Burt: "On at least one occasion he invented, for the purpose of quoting it in one of his articles, a thesis by one of his students never in fact written; at the time I interpreted this as a sign of forgetfulness."[25]

The Hearnshaw biography, published in 1979, has put to rest any lingering doubts about Burt's wholesale faking.[26] The painstaking searches and inquiries made by Hearnshaw failed to unearth any substantial traces of Miss Conway, or Miss Howard, or of any separated twins. There were many instances of dishonesty, of evasion, and of contradiction in Burt's written replies to correspondents who had inquired about his data. The evidence made clear that Burt had collected no data at all during the last thirty years of his life, when, supposedly, most of the separated twins had been studied. With painful reluctance, Hearnshaw found himself forced to conclude that the charges made by Burt's critics were "in their essentials valid." The evidence demonstrated that Burt had "fabricated figures" and had "falsified." There is now no doubt whatever that all of Burt's "data" on the heritability of IQ must be discarded. The loss of these incredibly clear-cut "data" has been devastating to the claim that a substantial IQ heritability was demonstrated.

But what are we to make of the additional fact that Burt's transparently fraudulent data were accepted for so long, and so uncritically, by the "experts" in the field? Perhaps the clearest moral to be drawn from the Burt affair was spelled out by N. J. Mackintosh in his review of the Hearnshaw biography in the *British Journal of Psychology:*

> Ignoring the question of fraud, the fact of the matter is that the crucial evidence that his data on IQ are scientifically unacceptable does not depend on any examination of Burt's diaries or correspondence. It is to be found in the data themselves. The evidence was there . . . in 1961. It was, indeed, clear to anyone with eyes to see in 1958. But it was not seen until 1972, when Kamin first pointed to Burt's totally inadequate reporting of his data and to the impossible consistencies in his correlation coefficients. Until then the data were cited, with respect bordering on reverence, as the most telling proof of the heritability of IQ. It is a sorry comment on the wider scientific community that "numbers . . . simply not worthy of our current scientific attention" . . . should have entered nearly every psychological textbook.[27]

We do not view the uncritical acceptance of Burt's data as an unusual or inexplicable "sorry comment on the wider scientific community." The fraud perpetrated by Burt, and unwittingly propagated by the scientific community, served important social purposes. Professor Hearnshaw's biography essentially saves the face of psychometry by probing the individual psychology of Burt to ask why he should have been moved to such fraudulence. Burt, no longer a nobleman but now victim of a debilitating and psychiatrically distressing disorder, has become the bad apple of psychometry. By 1980, when the British Psychological Society was prepared to draw up its "Balance Sheet on Burt,"[28] there had been a closing of the ranks; the psychometric doyens reiterated their belief that, despite the eviction of Burt, the residual evidence for the heritability of intelligence was strong. The social function of IQ ideology was still dominant.

NOTES

1. S. Bowles and V. Nelson, "The Inheritance of IQ and the Intergenerational Transmission of Economic Inequality." *Review of Economics and Statistics* 54, no. 1 (1974).
2. M. Schiff, M. Duyme, A. Dumaret, and S. Tomkiewicz, " 'How Much *Could* We Boost Scholastic Achievement and IQ Scores?' Direct Answer from a French Adoption Study." *Cognition* 12 (1982):165–96.
3. A. Binet. *Les Idées modernes sur les enfants* (Paris: Flammarion, 1913), pp. 140–41.
4. L. M. Terman, "Feeble-minded Children in the Public Schools of California." *School and Society* 5 (1917) 165.
5. L. M. Terman, *The Measurement of Intelligence* (Boston: Houghton Mifflin, 1916), pp. 91–92.
6. H. H. Goddard, *Human Efficiency and Levels of Intelligence* (Princeton, N.J.: Princeton University Press, 1920), pp. 99–103.
7. C. Burt, "Experimental Tests of General Intelligence." *British Journal of Psychology* 3 (1909):94–177.
8. C. Burt, *Mental and Scholastic Tests*, 2d ed. (London: Staples, 1947); and *The Backward Child*, 5th ed. (London: University of London Press, 1961).
9. L. Kamin, *The Science and Politics of IQ* (Potomac, Md.: Erlbaum, 1974); K. Ludmerer, *Genetics and American Society* (Baltimore: Johns Hopkins University Press, 1972); M. Haller, *Eugenics: Hereditarian Attitudes in American Thought* (New Brunswick, N.J.: Rutgers University Press, 1963); C. Karier, *The Making of the American Educational State* (Urbana: University of Illinois Press, 1973); and N. Stepan, *The Idea of Race in Science* (London: Macmillan, 1982).

10. E. G. Boring, "Intelligence as the Tests Test It." *New Republic* 34 (1923):35–36.
11. S. Bowles and V. Nelson, "The Inheritance of IQ and the Intergenerational Reproduction of Economic Inequality." *Review of Economics and Statistics* 56 (1974):39–51.
12. E. L. Thorndike, *Educational Psychology* (New York: Columbia University Teachers College, 1903) p. 140.
13. A. R. Jensen, "Sir Cyril Burt" (obituary), *Psychometrika* 37 (1972):115–17.
14. H. J. Eysenck, *The Inequality of Man* (London: Temple Smith, 1973).
15. C. Burt, "Ability and Income," *British Journal of Educational Psychology* 13 (1943):83–98.
16. C. Burt, "The Evidence for the Concept of Intelligence," *British Journal of Educational Psychology* 25 (1955):167–68.
17. H. J. Eysenck, "H. J. Eysenck in Rebuttal," *Change* 6, no. 2 (1974).
18. L. Kamin, "Heredity, Intelligence, Politics and Psychology," unpublished presidential address to meeting of the Eastern Psychological Association (1972).
19. Kamin, *Science and Politics of IQ*.
20. A. R. Jensen, "Kinship Correlations Reported by Sir Cyril Burt," *Behavior Genetics* 4 (1974):24–25.
21. A. R. Jensen, "How Much Can We Boost IQ and Scholastic Achievement?" *Howard Educational Review* 39 (1969):1–123.
22. O. Gillie, *Sunday Times* (London), 24 October 1976.
23. A. R. Jensen, "Heredity and Intelligence: Sir Cyril Burt's Findings," letters to the *Times* (London), 9 December 1976, p. 11.
24. H. J. Eysenck, "The Case of Sir Cyril Burt," *Encounter* 48 (1977):19–24.
25. H. J. Eysenck, "Sir Cyril Burt and the Inheritance of the IQ," *New Zealand Psychologist* (1978).
26. L. S. Hearnshaw, *Cyril Burt: Psychologist* (London: Hodder & Stoughton, 1979).
27. N. J. Mackintosh, book review of *Cyril Burt: Psychologist* by J. S. Hearnshaw, *British Journal of Psychology* 71 (1980):174–75.
28. A Balance Sheet on Cyril Burt, *Supplement to the Bulletin of the British Psychological Society* 33 (1980):i.