

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 Ménière's 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

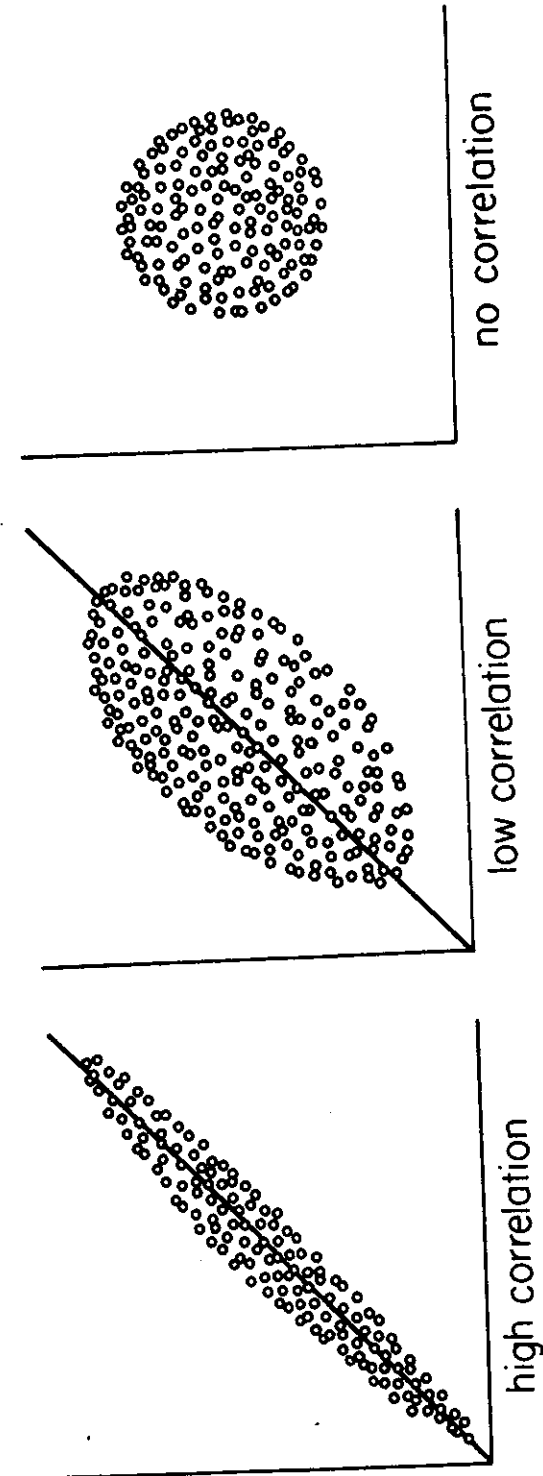
length and leg length from, say, age ten to adulthood would represent *zero correlation*—legs would get longer while teeth changed not at all. Other correlations can be negative—one measure increases while the other decreases. We begin to lose neurons at a distressingly early age, and they are not replaced. 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.



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.

tion to one dimension. Arm and leg length are tightly correlated because they are both partial measures of an underlying biological phenomenon, namely growth itself.

Yet, lest anyone become too hopeful that correlation represents a magic method for the unambiguous identification of cause, consider the relationship between my age and the price of gasoline during the past ten years. The correlation is nearly perfect, but no one would suggest any assignment of cause. The fact of correlation implies nothing about cause. It is not even true that intense correlations are more likely to represent cause than weak ones, for the correlation of my age with the price of gasoline is nearly 1.0. I spoke of cause for arm and leg lengths not because their correlation was high, but because I know something about the biology of the situation. The inference of cause must come from somewhere else, not from the simple fact of correlation—though an unexpected correlation may lead us to search for causes so long as we remember that we may not find them. The vast majority of correlations in our world are, without doubt, noncausal. Anything that has been increasing steadily during the past few years will be strongly correlated with the distance between the earth and Halley's comet (which has also been increasing of late)—but even the most dedicated astrologer would not discern causality in most of these relationships. The invalid assumption that correlation implies cause is probably among the two or three most serious and common errors of human reasoning.

Few people would be fooled by such a *reductio ad absurdum* as the age-gas correlation. But consider an intermediate case. I am given a table of data showing how far twenty children can hit and throw a baseball. I graph these data and calculate a high r . Most people, I think, would share my intuition that this is not a meaningless correlation; yet in the absence of further information, the correlation itself teaches me nothing about underlying causes. For I can suggest at least three different and reasonable causal interpretations for the correlation (and the true reason is probably some combination of them):

1. The children are simply of different ages, and older children can hit and throw farther.
2. The differences represent variation in practice and training. Some children are Little League stars and can tell you the year that

Rogers Hornsby hit .424 (1924—I was a bratty little kid like that); others know Billy Martin only as a figure in Lite beer commercials.

3. The differences represent disparities in native ability that cannot be erased even by intense training. (The situation would be even more complex if the sample included both boys and girls of conventional upbringing. The correlation might then be attributed primarily to a fourth cause—sexual differences; and we would have to worry, in addition, about the cause of the sexual difference: training, inborn constitution, or some combination of nature and nurture).

In summary, most correlations are noncausal; when correlations are causal, the fact and strength of the correlation rarely specifies the nature of the cause.

Correlation in more than two dimensions

These two-dimensional examples are easy to grasp (however difficult they are to interpret). But what of correlations among more than two measures? A body is composed of many parts, not just arms and legs, and we may want to know how several measures interact during growth. Suppose, for simplicity, that we add just one more measure, head length, to make a three-dimensional system. We may now depict the correlation structure among the three measures in two ways:

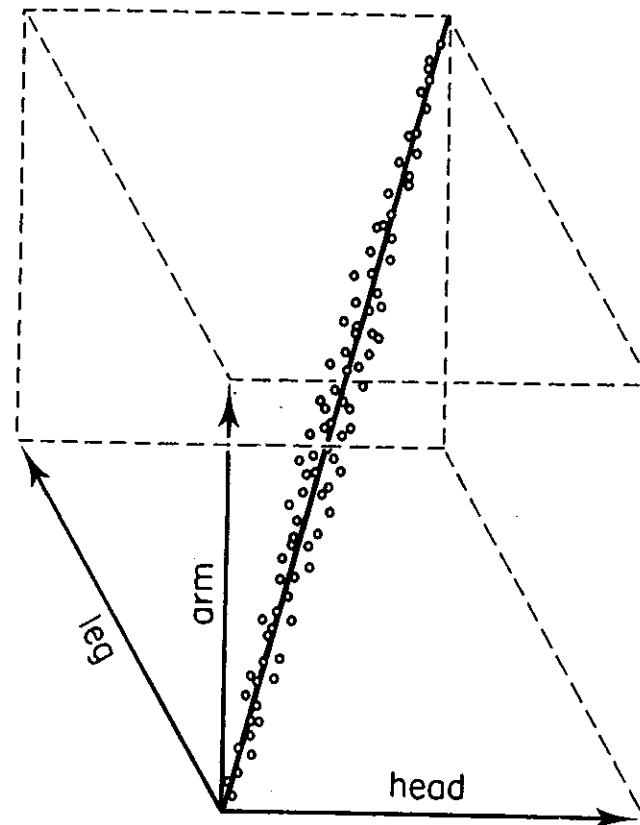
1. We may gather all correlation coefficients between pairs of measures into a single table, or *matrix* of correlation coefficients (Fig. 6.2). The line from upper left to lower right records the necessarily perfect correlation of each variable with itself. It is called the principal diagonal, and all correlations along it are 1.0. The matrix is symmetrical around the principal diagonal, since the correlation of measure 1 with measure 2 is the same as the correlation of 2 with 1. Thus, the three values either above or below the principal diagonal are the correlations we seek: arm with leg, arm with head, and leg with head.

2. We may plot the points for all individuals onto a three-dimensional graph (Fig. 6.3). Since the correlations are all positive, the points are oriented as an ellipsoid (or football). (In two dimensions, they formed an ellipse.) A line running along the major axis of the football expresses the strong positive correlations between all measures.

	arm	leg	head
arm	1.0	0.91	0.72
leg	0.91	1.0	0.63
head	0.72	0.63	1.0

6•2 A correlation matrix for three measurements.

6•3 A three-dimensional graph showing the correlations for three measurements.



We can grasp the three-dimensional case, both mentally and pictorially. But what about 20 dimensions, or 100? If we measured 100 parts of a growing body, our correlation matrix would contain 10,000 items. To plot this information, we would have to work in a 100-dimensional space, with 100 mutually perpendicular axes representing the original measures. Although these 100 axes present no mathematical problem (they form, in technical terms, a hyperspace), we cannot plot them in our three-dimensional Euclidean world.

These 100 measures of a growing body probably do not represent 100 different biological phenomena. Just as most of the information in our three-dimensional example could be resolved into a single dimension (the long axis of the football), so might our 100 measures be simplified into fewer dimensions. We will lose some information in the process to be sure—as we did when we collapsed the long and skinny football, still a three-dimensional structure, into the single line representing its long axis. But we may be willing to accept this loss in exchange for simplification and for the possibility of interpreting the dimensions that we do retain in biological terms.

Factor analysis and its goals

With this example, we come to the heart of what *factor analysis* attempts to do. Factor analysis is a mathematical technique for reducing a complex system of correlations into fewer dimensions. It works, literally, by factoring a matrix, usually a matrix of correlation coefficients. (Remember the high-school algebra exercise called "factoring," where you simplified horrendous expressions by removing common multipliers of all terms?) Geometrically, the process of factoring amounts to placing axes through a football of points. In the 100-dimensional case, we are not likely to recover enough information on a single line down the hyperfootball's long axis—a line called the *first principal component*. We will need additional axes. By convention, we represent the second dimension by a line *perpendicular* to the first principal component. This second axis, or *second principal component*, is defined as the line that resolves more of the remaining variation than any other line that could be drawn perpendicular to the first principal component. If, for example, the hyperfootball were squashed flat like a flounder, the

first principal component would run through the middle, from head to tail, and the second also through the middle, but from side to side. Subsequent lines would be perpendicular to all previous axes, and would resolve a steadily decreasing amount of remaining variation. We might find that five principal components resolve almost all the variation in our hyperfootball—that is, the hyperfootball drawn in 5 dimensions looks sufficiently like the original to satisfy us, just as a pizza or a flounder drawn in two dimensions may express all the information we need, even though both original objects contain three dimensions. If we elect to stop at 5 dimensions, we may achieve a considerable simplification at the acceptable price of minimal loss of information. We can grasp the 5 dimensions conceptually; we may even be able to interpret them biologically.

Since factoring is performed on a correlation matrix, I shall use a geometrical representation of the correlation coefficients themselves in order to explain better how the technique operates. The original measures may be represented as vectors of unit length,*

*(Footnote for aficionados—others may safely skip.) Here, I am technically discussing a procedure called "principal components analysis," not quite the same thing as factor analysis. In principal components analysis, we preserve all information in the original measures and fit new axes to them by the same criterion used in factor analysis in principal components orientation—that is, the first axis explains more data than any other axis could and subsequent axes lie at right angles to all other axes and encompass steadily decreasing amounts of information. In true factor analysis, we decide beforehand (by various procedures) not to include all information on our factor axes. But the two techniques—true factor analysis in principal components orientation and principal components analysis—play the same conceptual role and differ only in mode of calculation. In both, the first axis (Spearman's *g* for intelligence tests) is a "best fit" dimension that resolves more information in a set of vectors than any other axis could.

During the past decade or so, semantic confusion has spread in statistical circles through a tendency to restrict the term "factor analysis" only to the rotations of axes usually performed after the calculation of principal components, and to extend the term "principal components analysis" both to true principal components analysis (all information retained) and to factor analysis done in principal components orientation (reduced dimensionality and loss of information). This shift in definition is completely out of keeping with the history of the subject and terms. Spearman, Burt, and hosts of other psychometricians worked for decades in this area before Thurstone and others invented axial rotations. They performed all their calculations in the principal components orientation, and they called themselves "factor analysts." I continue, therefore, to use the term "factor analysis" in its original sense to include any orientation of axes—principal components or rotated, orthogonal or oblique.

I will also use a common, if somewhat sloppy, shorthand in discussing what

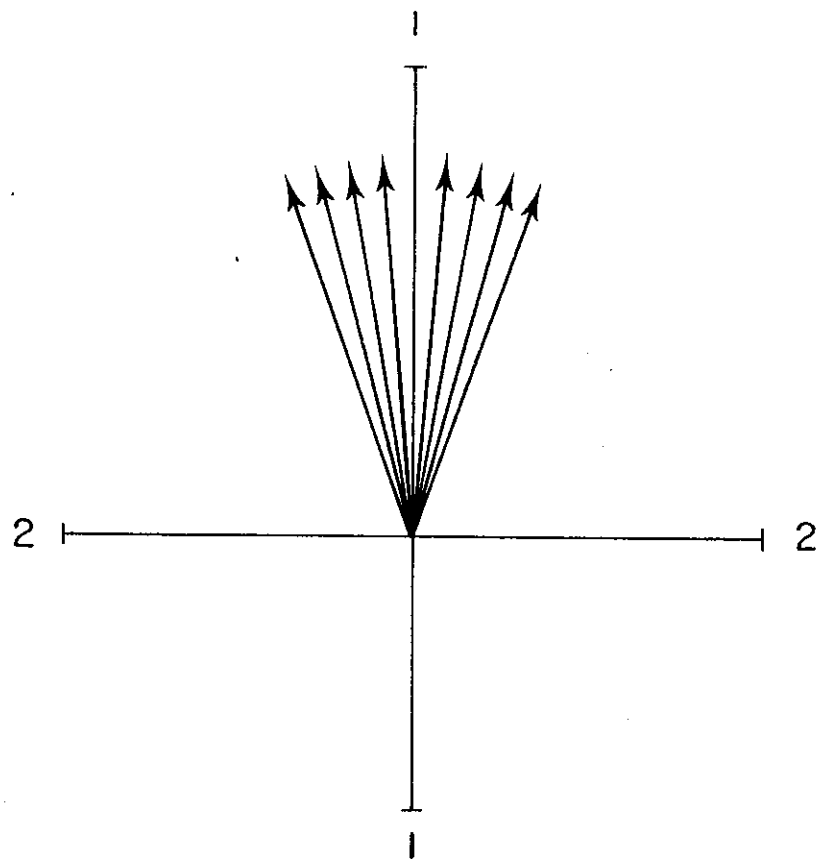
radiating from a common point. If two measures are highly correlated, their vectors lie close to each other. The cosine of the angle between any two vectors records the correlation coefficient between them. If two vectors overlap, their correlation is perfect, or 1.0; the cosine of 0° is 1.0. If two vectors lie at right angles, they are completely independent, with a correlation of zero; the cosine of 90° is zero. If two vectors point in opposite directions, their correlation is perfectly negative, or -1.0 ; the cosine of 180° is -1.0 . A matrix of high positive correlation coefficients will be represented by a cluster of vectors, each separated from each other vector by a small acute angle (Fig. 6.4). When we factor such a cluster into fewer dimensions by computing principal components, we choose as our first component the axis of maximal resolving power, a kind of grand average among all vectors. We assess resolving power by projecting each vector onto the axis. This is done by drawing a line from the tip of the vector to the axis, perpendicular to the axis. The ratio of projected length on the axis to the actual length of the vector itself measures the percentage of a vector's information resolved by the axis. (This is difficult to express verbally, but I think that Figure 6.5 will dispel confusion.) If a vector lies near the axis, it is highly resolved and the axis encompasses most of its information. As a vector moves away from the axis toward a maximal separation of 90° , the axis resolves less and less of it.

We position the first principal component (or axis) so that it resolves more information among all the vectors than any other axis could. For our matrix of high positive correlation coefficients, represented by a set of tightly clustered vectors, the first principal component runs through the middle of the set (Fig. 6.4). The second principal component lies at right angles to the first and resolves a maximal amount of remaining information. But if the first component has already resolved most of the information in all the vectors, then the second and subsequent principal axes can only deal with the small amount of information that remains (Fig. 6.4).

factor axes do. Technically, factor axes resolve variance in original measures. I will, as is often done, speak of them as "explaining" or "resolving" information—as they do in the vernacular (though not in the technical) sense of information. That is, when the vector of an original variable projects strongly on a set of factor axes, little of its variance lies unresolved in higher dimensions outside the system of factor axes.

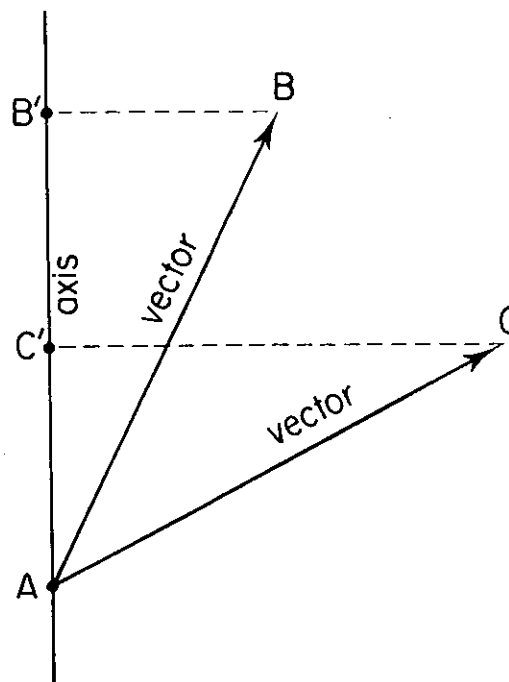
Such systems of high positive correlation are found frequently in nature. In my own first study in factor analysis, for example, I considered fourteen measurements on the bones of twenty-two species of pelycosaurian reptiles (the fossil beasts with the sails on their backs, often confused with dinosaurs, but actually the ancestors of mammals). My first principal component resolved 97.1 per-

6•4 Geometric representation of correlations among eight tests when all correlation coefficients are high and positive. The first principal component, labeled 1, lies close to all the vectors, while the second principal component, labeled 2, lies at right angles to the first and does not explain much information in the vectors.



cent of the information in all fourteen vectors, leaving only 2.9 percent for subsequent axes. My fourteen vectors formed an extremely tight swarm (all practically overlapping); the first axis went through the middle of the swarm. My pelycosaurs ranged in body length from less than two to more than eleven feet. They all look pretty much alike, and big animals have larger measures for all fourteen bones. All correlation coefficients of bones with other bones are very high; in fact, the lowest is still a whopping 0.912.

6•5 Computing the amount of information in a vector explained by an axis. Draw a line from the tip of the vector to the axis, perpendicular to the axis. The amount of information resolved by the axis is the ratio of the projected length on the axis to the true length of the vector. If a vector lies close to the axis, then this ratio is high and most of the information in the vector is resolved by the axis. Vector AB lies close to the axis and the ratio of the projection AB' to the vector itself, AB, is high. Vector AC lies far from the axis and the ratio of its projected length AC' to the vector itself, AC, is low.



Scarcely surprising. After all, large animals have large bones, and small animals small bones. I can interpret my first principal component as an abstracted size factor, thus reducing (with minimal loss of information) my fourteen original measurements into a single dimension interpreted as increasing body size. In this case, factor analysis has achieved both *simplification* by reduction of dimensions (from fourteen to effectively one), and *explanation* by reasonable biological interpretation of the first axis as a size factor.

But—and here comes an enormous but—before we rejoice and extol factor analysis as a panacea for understanding complex systems of correlation, we should recognize that it is subject to the same cautions and objections previously examined for the correlation coefficients themselves. I consider two major problems in the following sections.

The error of reification

The first principal component is a mathematical abstraction that can be calculated for any matrix of correlation coefficients; it is not a "thing" with physical reality. Factorists have often fallen prey to a temptation for *reification*—for awarding *physical meaning* to all strong principal components. Sometimes this is justified; I believe that I can make a good case for interpreting my first pelycosaurian axis as a size factor. But such a claim can never arise from the mathematics alone, only from additional knowledge of the physical nature of the measures themselves. For nonsensical systems of correlation have principal components as well, and they may resolve more information than meaningful components do in other systems. A factor analysis for a five-by-five correlation matrix of my age, the population of Mexico, the price of swiss cheese, my pet turtle's weight, and the average distance between galaxies during the past ten years will yield a strong first principal component. This component—since all the correlations are so strongly positive—will probably resolve as high a percentage of information as the first axis in my study of pelycosaur. It will also have no enlightening physical meaning whatever.

In studies of intelligence, factor analysis has been applied to matrices of correlation among mental tests. Ten tests may, for example, be given to each of one hundred people. Each meaningful entry in the ten-by-ten correlation matrix is a correlation coef-

ficient between scores on two tests taken by each of the one hundred persons. We have known since the early days of mental testing—and it should surprise no one—that most of these correlation coefficients are positive: that is, people who score highly on one kind of test tend, on average, to score highly on others as well. Most correlation matrices for mental tests contain a preponderance of positive entries. This basic observation served as the starting point for factor analysis. Charles Spearman virtually invented the technique in 1904 as a device for inferring causes from correlation matrices of mental tests.

Since most correlation coefficients in the matrix are positive, factor analysis must yield a reasonably strong first principal component. Spearman calculated such a component indirectly in 1904 and then made the cardinal invalid inference that has plagued factor analysis ever since. He reified it as an "entity" and tried to give it an unambiguous causal interpretation. He called it *g*, or general intelligence, and imagined that he had identified a unitary quality underlying all cognitive mental activity—a quality that could be expressed as a single number and used to rank people on a unilinear scale of intellectual worth.

Spearman's *g*—the first principal component of the correlation matrix of mental tests—never attains the predominant role that a first component plays in many growth studies (as in my pelycosaur). At best, *g* resolves 50 to 60 percent of all information in the matrix of tests. Correlations between tests are usually far weaker than correlations between two parts of a growing body. In most cases, the highest correlation in a matrix of tests does not come close to reaching the *lowest* value in my pelycosaur matrix—0.912.

Although *g* never matches the strength of a first principal component of some growth studies, I do not regard its fair resolving power as accidental. Causal reasons lie behind the positive correlations of most mental tests. But what reasons? We cannot infer the reasons from a strong first principal component any more than we can induce the cause of a single correlation coefficient from its magnitude. We cannot reify *g* as a "thing" unless we have convincing, independent information beyond the fact of correlation itself.

The situation for mental tests resembles the hypothetical case I presented earlier of correlation between throwing and hitting a baseball. The relationship is strong and we have a right to regard

it as nonaccidental. But we cannot infer the cause from the correlation, and the cause is certainly complex.

Spearman's g is particularly subject to ambiguity in interpretation, if only because the two most contradictory causal hypotheses are both fully consistent with it: 1) that it reflects an inherited level of mental acuity (some people do well on most tests because they are born smarter); or 2) that it records environmental advantages and deficits (some people do well on most tests because they are well schooled, grew up with enough to eat, books in the home, and loving parents). If the simple existence of g can be theoretically interpreted in either a purely hereditarian or purely environmentalist way, then its mere presence—even its reasonable strength—cannot justly lead to any reification at all. The temptation to reify is powerful. The idea that we have detected something “underlying” the externalities of a large set of correlation coefficients, something perhaps more real than the superficial measurements themselves, can be intoxicating. It is Plato's essence, the abstract, eternal reality underlying superficial appearances. But it is a temptation that we must resist, for it reflects an ancient prejudice of thought, not a truth of nature.

Rotation and the nonnecessity of principal components

Another, more technical, argument clearly demonstrates why principal components cannot be automatically reified as causal entities. If principal components represented the only way to simplify a correlation matrix, then some special status for them might be legitimately sought. But they represent only one method among many for inserting axes into a multidimensional space. Principal components have a definite geometric arrangement, specified by the criterion used to construct them—that the first principal component shall resolve a maximal amount of information in a set of vectors and that subsequent components shall all be mutually perpendicular. But there is nothing sacrosanct about this criterion; vectors may be resolved into any set of axes placed within their space. Principal components provide insight in some cases, but other criteria are often more useful.

Consider the following situation, in which another scheme for placing axes might be preferred. In Figure 6.6 I show correlations between four mental tests, two of verbal and two of arithmetical

aptitude. Two “clusters” are evident, even though all tests are positively correlated. Suppose that we wish to identify these clusters by factor analysis. If we use principal components, we may not recognize them at all. The first principal component (Spearman's g) goes right up the middle, between the two clusters. It lies close to no vector and resolves an approximately equal amount of each, thereby masking the existence of verbal and arithmetic clusters. Is this component an entity? Does a “general intelligence” exist? Or is g , in this case, merely a meaningless average based on the invalid amalgamation of two types of information?

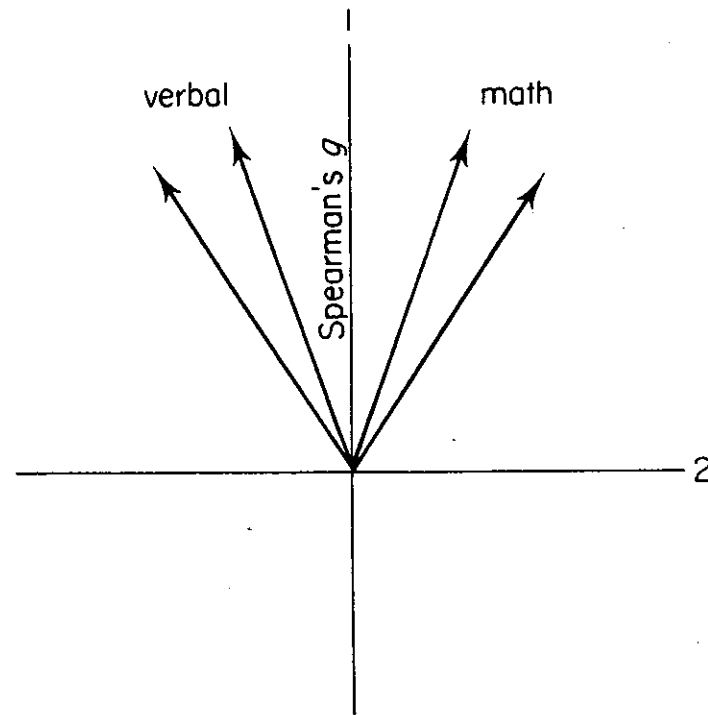
We may pick up verbal and arithmetic clusters on the second principal component (called a “bipolar factor” because some projections upon it will be positive and others negative when vectors lie on both sides of the first principal component). In this case, verbal tests project on the negative side of the second component, and arithmetic tests on the positive side. But we may fail to detect these clusters altogether if the first principal component dominates all vectors. For projections on the second component will then be small, and the pattern can easily be lost (see Fig. 6.6).

During the 1930s factorists developed methods to treat this dilemma and to recognize clusters of vectors that principal components often obscured. They did this by rotating factor axes from the principal components orientation to new positions. The rotations, established by several criteria, had as their common aim the positioning of axes near clusters. In Figure 6.7, for example, we use the criterion: place axes near vectors occupying extreme or outlying positions in the total set. If we now resolve all vectors into these rotated axes, we detect the clusters easily; for arithmetic tests project high on rotated axis 1 and low on rotated axis 2, while verbal tests project high on 2 and low on 1. Moreover, g has disappeared. We no longer find a “general factor” of intelligence, nothing that can be reified as a single number expressing overall ability. Yet we have lost no information. The two rotated axes resolve as much information in the four vectors as did the two principal components. They simply distribute the same information differently upon the resolving axes. How can we argue that g has any claim to reified status as an entity if it represents but one of numerous possible ways to position axes within a set of vectors?

In short, factor analysis simplifies large sets of data by reducing

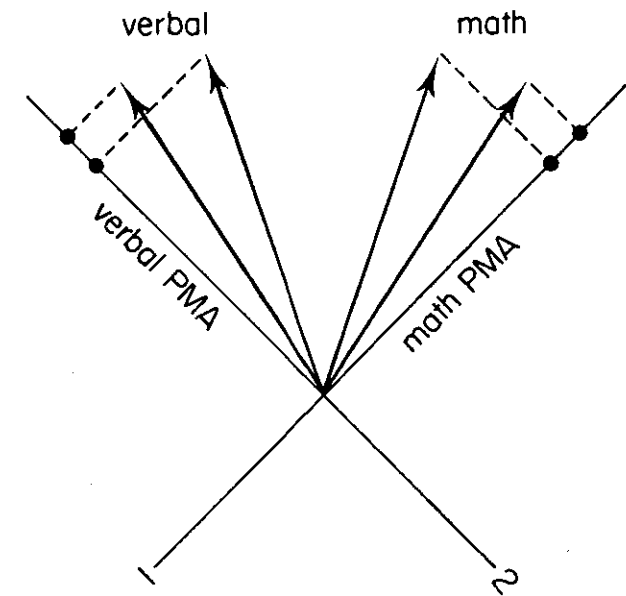
dimensionality and trading some loss of information for the recognition of ordered structure in fewer dimensions. As a tool for simplification, it has proved its great value in many disciplines. But many factorists have gone beyond simplification, and tried to define factors as causal entities. This error of reification has plagued the technique since its inception. It was "present at the creation" since Spearman invented factor analysis to study the correlation matrix of mental tests and then reified his principal component as *g* or innate, general intelligence. Factor analysis may help us to understand causes by directing us to information beyond the

6•6 A principal components analysis of four mental tests. All correlations are high and the first principal component, Spearman's *g*, expresses the overall correlation. But the group factors for verbal and mathematical aptitude are not well resolved in this style of analysis.



mathematics of correlation. But factors, by themselves, are neither things nor causes; they are mathematical abstractions. Since the same set of vectors (see Figs. 6.6, 6.7) can be partitioned into *g* and a small residual axis, or into two axes of equal strength that identify verbal and arithmetical clusters and dispense with *g* entirely, we cannot claim that Spearman's "general intelligence" is an ineluctable entity necessarily underlying and causing the correlations among mental tests. Even if we choose to defend *g* as a nonaccidental result, neither its strength nor its geometric position can specify what it means in causal terms—if only because its features are equally consistent with extreme hereditarian and extreme environmentalist views of intelligence.

6•7 Rotated factor axes for the same four mental tests depicted in Fig. 6.6. Axes are now placed near vectors lying at the periphery of the cluster. The group factors for verbal and mathematical aptitude are now well identified (see high projections on the axes indicated by dots), but *g* has disappeared.



Charles Spearman and general intelligence

The two-factor theory

Correlation coefficients are now about as ubiquitous and unsurprising as cockroaches in New York City. Even the cheapest pocket calculators produce correlation coefficients with the press of a button. However indispensable, they are taken for granted as automatic accouterments of any statistical analysis that deals with more than one measure. In such a context, we easily forget that they were once hailed as a breakthrough in research, as a new and exciting tool for discovering underlying structure in tables of raw measures. We can sense this excitement in reading early papers of the great American biologist and statistician Raymond Pearl (see Pearl, 1905 and 1906, and Pearl and Fuller, 1905). Pearl completed his doctorate at the turn of the century and then proceeded, like a happy boy with a gleaming new toy, to correlate everything in sight, from the lengths of earth worms vs. the number of their body segments (where he found no correlation and assumed that increasing length reflects larger, rather than more, segments), to size of the human head vs. intelligence (where he found a very small correlation, but attributed it to the indirect effect of better nutrition).

Charles Spearman, an eminent psychologist and fine statistician as well* began to study correlations between mental tests during these heady times. If two mental tests are given to a large number of people, Spearman noted, the correlation coefficient between them is nearly always positive. Spearman pondered this result and wondered what higher generality it implied. The positive correlations clearly indicated that each test did not measure an independent attribute of mental functioning. Some simpler structure lay behind the pervasive positive correlations; but what structure? Spearman imagined two alternatives. First, the positive correlations might reduce to a small set of independent attributes—the “faculties” of the phrenologists and other schools of early psychology. Perhaps the mind had separate “compartments” for arithmetic, verbal, and spatial aptitudes, for example. Spearman called such

*Spearman took a special interest in problems of correlation and invented a measure that probably ranks second in use to Pearson's r as a measure of association between two variables—the so-called Spearman's rank-correlation coefficient.

theories of intelligence “oligarchic.” Second, the positive correlations might reduce to a single, underlying general factor—a notion that Spearman called “monarchic.” In either case, Spearman recognized that the underlying factors—be they few (oligarchic) or single (monarchic)—would not encompass all information in a matrix of positive correlation coefficients for a large number of mental tests. A “residual variance” would remain—information peculiar to each test and not related to any other. In other words, each test would have its “anarchic” component. Spearman called the residual variance of each test its s , or specific information. Thus, Spearman reasoned, a study of underlying structure might lead to a “two-factor theory” in which each test contained some specific information (its s) and also reflected the operation of a single, underlying factor, which Spearman called g , or general intelligence. Or each test might include its specific information and also record one or several among a set of independent, underlying faculties—a many-factor theory. If the simplest two-factor theory held, then all common attributes of intelligence would reduce to a single underlying entity—a true “general intelligence” that might be measured for each person and might afford an unambiguous criterion for ranking in terms of mental worth.

Charles Spearman developed factor analysis—still the most important technique in modern multivariate statistics—as a procedure for deciding between the two- vs. the many-factor theory by determining whether the common variance in a matrix of correlation coefficients could be reduced to a single “general” factor, or only to several independent “group” factors. He found but a single “intelligence,” opted for the two-factor theory, and, in 1904, published a paper that later won this assessment from a man who opposed its major result: “No single event in the history of mental testing has proved to be of such momentous importance as Spearman's proposal of his famous two-factor theory” (Guilford, 1936, p. 155). Elated, and with characteristic immodesty, Spearman gave his 1904 paper a heroic title: “General Intelligence Objectively Measured and Determined.” Ten years later (1914, p. 237), he exulted: “The future of research into the inheritance of ability must center on the theory of ‘two factors.’ This alone seems capable of reducing the bewildering chaos of facts to a perspicuous orderliness. By its means, the problems are rendered clear; in many

respects, their answers are already foreshadowed; and everywhere, they are rendered susceptible of eventual decisive solution."

The method of tetrad differences

In his original work, Spearman did not use the method of principal components described on pp. 275–278. Instead, he developed a simpler, though tedious, procedure better suited for a precomputer age when all calculations had to be performed by hand.* He computed the entire matrix of correlation coefficients between all pairs of tests, took all possible groupings of four measures and computed for each a number that he called the "tetrad difference." Consider the following example as an attempt to define the tetrad difference and to explain how Spearman used it to test whether the common variance of his matrix could be reduced to a single general factor, or only to several group factors.

Suppose that we wish to compute the tetrad difference for four measures taken on a series of mice ranging in age from babies to adults—leg length, leg width, tail length, and tail width. We compute all correlation coefficients between pairs of variables and find, unsurprisingly, that all are positive—as mice grow, their parts get larger. But we would like to know whether the common variance in the positive correlations all reflects a single general factor—growth itself—or whether two separate components of growth must be identified—in this case, a leg factor and a tail factor, or a length factor and a width factor. Spearman gives the following formula for the tetrad difference

$$r_{13} \times r_{24} - r_{23} \times r_{14}$$

where r is the correlation coefficient and the two subscripts represent the two measures being correlated (in this case, 1 is leg length, 2 is leg width, 3 is tail length and 4 is tail width—so that r_{13} is the correlation coefficient between the first and the third measure, or between leg length and tail length). In our example, the tetrad difference is

$$\begin{aligned} &(\text{leg length and tail length}) \times (\text{leg width and tail width}) - \\ &(\text{leg width and tail length}) \times (\text{leg length and tail width}). \end{aligned}$$

*The g calculated by the tetrad formula is conceptually equivalent and mathematically almost equivalent to the first principal component described on pp. 275–278 and used in modern factor analysis.

Spearman argued that tetrad differences of zero imply the existence of a single general factor while either positive or negative values indicate that group factors must be recognized. Suppose, for example, that group factors for general body length and general body width govern the growth of mice. In this case, we would get a high positive value for the tetrad difference because the correlation coefficients of a length with another length or a width with another width would tend to be higher than correlation coefficients of a width with a length. (Note that the left-hand side of the tetrad equation includes only lengths with lengths or widths with widths, while the right-hand side includes only lengths with widths.) But if only a single, general growth factor regulates the size of mice, then lengths with widths should show as high a correlation as lengths with lengths or widths with widths—and the tetrad difference should be zero. Fig. 6.8 shows a hypothetical correlation matrix for the four measures that yields a tetrad difference of zero (values taken from Spearman's example in another context, 1927, p. 74). Fig. 6.8 also shows a different hypothetical matrix yielding a positive tetrad difference and a conclusion (if other tetrads show the same pattern) that group factors for length and width must be recognized.

The top matrix of Fig. 6.8 illustrates another important point that reverberates throughout the history of factor analysis in psychology. Note that, although the tetrad difference is zero, the correlation coefficients need not be (and almost invariably are not) equal. In this case, leg width with leg length gives a correlation of 0.80, while tail width with tail length yields only 0.18. These differences reflect varying "saturation" with g , the single general factor when the tetrad differences are zero. Leg measures have higher saturations than tail measures—that is, they are closer to g , or reflect it better (in modern terms, they lie closer to the first principal component in geometric representations like Fig. 6.6). Tail measures do not load strongly on g .^{*} They contain little common variance and must be explained primarily by their s —the information unique to each measure. Moving now to mental tests: if g represents general intelligence, then mental tests most saturated with

*The terms "saturation" and "loading" refer to the correlation between a test and a factor axis. If a test "loads" strongly on a factor then most of its information is explained by the factor.

	LL	LW	TL	TW
LL	1.0			
LW	0.80	1.0		
TL	0.60	0.48	1.0	
TW	0.30	0.24	0.18	1.0

Tetrad difference:
 $0.60 \times 0.24 - 0.48 \times 0.30$
 $0.144 - 0.144 = 0$
 no group factors

	LL	LW	TL	TW
LL	1.0			
LW	0.80	1.0		
TL	0.40	0.20	1.0	
TW	0.20	0.40	0.50	1.0

Tetrad difference:
 $0.40 \times 0.40 - 0.20 \times 0.20$
 $0.16 - 0.04 = 0.12$
 group factors for lengths
 and widths

6•8 Tetrad differences of zero (above) and a positive value (below) from hypothetical correlation matrices for four measurements: LL = leg length, LW = leg width, TL = tail length, and TW = tail width. The positive tetrad difference indicates the existence of group factors for lengths and widths.

g are the best surrogates for general intelligence, while tests with low g -loadings (and high s values) cannot serve as good measures of general mental worth. Strength of g -loading becomes the criterion for determining whether or not a particular mental test (IQ, for example) is a good measure of general intelligence.

Spearman's tetrad procedure is very laborious when the correlation matrix includes a large number of tests. Each tetrad difference must be calculated separately. If the common variance reflects but a single general factor, then the tetrads should equal zero. But, as in any statistical procedure, not all cases meet the expected value (half heads and half tails is the expectation in coin flipping, but you will flip six heads in a row about once in sixty-four series of six flips). Some calculated tetrad differences will be positive or negative even when a single g exists and the expected value is zero. Thus, Spearman computed all tetrad differences and looked for normal frequency distributions with a mean tetrad difference of zero as his test for the existence of g .

Spearman's g and the great instauration of psychology

Charles Spearman computed all his tetrads, found a distribution close enough to normal with a mean close enough to zero, and proclaimed that the common variance in mental tests recorded but a single underlying factor—Spearman's g , or general intelligence. Spearman did not hide his pleasure, for he felt that he had discovered the elusive entity that would make psychology a true science. He had found the innate essence of intelligence, the reality underlying all the superficial and inadequate measures devised to search for it. Spearman's g would be the philosopher's stone of psychology, its hard, quantifiable "thing"—a fundamental particle that would pave the way for an exact science as firm and as basic as physics.

In his 1904 paper, Spearman proclaimed the ubiquity of g in all processes deemed intellectual: "All branches of intellectual activity have in common one fundamental function . . . whereas the remaining or specific elements seem in every case to be wholly different from that in all the others. . . . This g , far from being confined to some small set of abilities whose intercorrelations have actually been measured and drawn up in some particular table, may enter into all abilities whatsoever."

The conventional school subjects, insofar as they reflect aptitude rather than the simple acquisition of information, merely peer through a dark glass at the single essence inside: "All examination in the different sensory, school, and other specific faculties may be considered as so many independently obtained estimates of the one great common Intellectual Function" (1904, p. 273). Thus Spearman tried to resolve a traditional dilemma of conventional education for the British elite: why should training in the classics make a better soldier or a statesman? "Instead of continuing ineffectively to protest that high marks in Greek syntax are no test as to the capacity of men to command troops or to administer provinces, we shall at last actually determine the precise accuracy of the various means of measuring General Intelligence" (1904, p. 277). In place of fruitless argument, one has simply to determine the *g*-loading of Latin grammar and military acuity. If both lie close to *g*, then skill in conjugation may be a good estimate of future ability to command.

There are different styles of doing science, all legitimate and partially valid. The beetle taxonomist who delights in noting the peculiarities of each new species may have little interest in reduction, synthesis, or in probing for the essence of "beetleness"—if such exists! At an opposite extreme, occupied by Spearman, the externalities of this world are only superficial guides to a simpler, underlying reality. In a popular image (though some professionals would abjure it), physics is the ultimate science of reduction to basic and quantifiable causes that generate the apparent complexity of our material world. Reductionists like Spearman, who work in the so-called soft sciences of organismic biology, psychology, or sociology, have often suffered from "physics envy." They have strived to practice their science according to their clouded vision of physics—to search for simplifying laws and basic particles. Spearman described his deepest hopes for a science of cognition (1923, p. 30):

Deeper than the uniformities of occurrence which are noticeable even without its aid, it [science] discovers others more abstruse, but correspondingly more comprehensive, upon which the name of laws is bestowed. . . . When we look around for any approach to this ideal, something of the sort can actually be found in the science of physics as based on the three primary laws of motion. Coordinate with this *physica corporis* [physics of bodies], then, we are today in search of a *physica animae* [physics of the soul].

With *g* as a quantified, fundamental particle, psychology could take its rightful place among the real sciences. "In these principles," he wrote in 1923 (p. 355), "we must venture to hope that the so long missing genuinely scientific foundation for psychology has at last been supplied, so that it can henceforward take its due place along with the other solidly founded sciences, even physics itself." Spearman called his work "a Copernican revolution in point of view" (1927, p. 411) and rejoiced that "this Cinderella among the sciences has made a bold bid for the level of triumphant physics itself" (1937, p. 21).

Spearman's g and the theoretical justification of IQ

Spearman, the theorist, the searcher for unity by reduction to underlying causes, often spoke in most unflattering terms about the stated intentions of IQ testers. He referred to IQ (1931) as "the mere average of sub-tests picked up and put together without rhyme or reason." He decried the dignification of this "gallimaufry of tests" with the name intelligence. In fact, though he had described his *g* as general intelligence in 1904, he later abandoned the word intelligence because endless arguments and inconsistent procedures of mental testers had plunged it into irremediable ambiguity (1927, p. 412; 1950, p. 67).

Yet it would be incorrect—indeed it would be precisely contrary to Spearman's view—to regard him as an opponent of IQ testing. He had contempt for the atheoretical empiricism of the testers, their tendency to construct tests by throwing apparently unrelated items together and then offering no justification for such a curious procedure beyond the claim that it yielded good results. Yet he did not deny that the Binet tests worked, and he rejoiced in the resuscitation of the subject thus produced: "By this one great investigation [the Binet scale] the whole scene was transformed. The recently despised tests were now introduced into every country with enthusiasm. And everywhere their practical application was brilliantly successful" (1914, p. 312).

What galled Spearman was his conviction that IQ testers were doing the right thing in amalgamating an array of disparate items into a single scale, but that they refused to recognize the theory behind such a procedure and continued to regard their work as rough-and-ready empiricism.

Spearman argued passionately that the justification for Binet

testing lay with his own theory of a single g underlying all cognitive activity. IQ tests worked because, unbeknownst to their makers, they measured g with fair accuracy. Each individual test has a g -loading and its own specific information (or s), but g -loading varies from nearly zero to nearly 100 percent. Ironically, the most accurate measure of g will be the average score for a large collection of individual tests of the most diverse kind. Each measures g to some extent. The variety guarantees that s -factors of the individual tests will vary in all possible directions and cancel each other out. Only g will be left as the factor common to all tests. IQ works because it measures g .

An explanation is at once supplied for the success of their extraordinary procedure of . . . pooling together tests of the most miscellaneous description. For if every performance depends on two factors, the one always varying randomly, while the other is constantly the same, it is clear that in the average the random variations will tend to neutralize one another, leaving the other, or constant factor, alone dominant (1914, p. 313; see also, 1923, p. 6, and 1927, p. 77).

Binet's "hotchpot of multitudinous measurements" was a correct theoretical decision, not only the intuitive guess of a skilled practitioner: "In such wise this principle of making a hotchpot, which might seem to be the most arbitrary and meaningless procedure imaginable, had really a profound theoretical basis and a supremely practical utility" (Spearman quoted in Tuddenham, 1962, p. 503).

Spearman's g , and its attendant claim that intelligence is a single, measurable entity, provided the only promising theoretical justification that hereditarian theories of IQ have ever had. As mental testing rose to prominence during the early twentieth century, it developed two traditions of research that Cyril Burt correctly identified in 1914 (p. 36) as correlational methods (factor analysis) and age-scale methods (IQ testing). Hearnshaw has recently made the same point in his biography of Burt (1979, p. 47): "The novelty of the 1900's was not in the concept of intelligence itself, but in its operational definition in terms of correlational techniques, and in the devising of practicable methods of measurement."

No one recognized better than Spearman the intimate connection between his model of factor analysis and hereditarian interpretations of IQ testing. In his 1914 *Eugenics Review* article, he

prophesied the union of these two great traditions in mental testing: "Each of these two lines of investigation furnishes a peculiarly happy and indispensable support to the other. . . . Great as has been the value of the Simon-Binet tests, even when worked in theoretical darkness, their efficiency will be multiplied a thousand-fold when employed with a full light upon their essential nature and mechanism." When Spearman's style of factor analysis came under attack late in his career (see pp. 326-332), he defended g by citing it as the rationale for IQ: "Statistically, this determination is grounded on its extreme simpleness. Psychologically, it is credited with affording the sole base for such useful concepts as those of 'general ability,' or 'IQ'" (1939, p. 79).

To be sure, the professional testers did not always heed Spearman's plea for an adoption of g as the rationale for their work. Many testers abjured theory and continued to insist on practical utility as the justification for their efforts. But silence about theory does not connote an absence of theory. The reification of IQ as a biological entity has depended upon the conviction that Spearman's g measures a single, scalable, fundamental "thing" residing in the human brain. Many of the more theoretically inclined mental testers have taken this view (see Terman et al., 1917, p. 152). C. C. Brigham did not base his famous recantation solely upon a belated recognition that the army mental tests had considered patent measures of culture as inborn properties (pp. 262-263). He also pointed out that no strong, single g could be extracted from the combined tests, which, therefore, could not have been measures of intelligence after all (Brigham, 1930). And I will at least say this for Arthur Jensen: he recognizes that his hereditarian theory of IQ depends upon the validity of g , and he devotes much of his major book (1979) to a defense of Spearman's argument in its original form, as do Richard Herrnstein and Charles Murray in *The Bell Curve* (1994)—see essays at end of this book. A proper understanding of the conceptual errors in Spearman's formulation is a prerequisite for criticizing hereditarian claims about IQ at their fundamental level, not merely in the tangled minutiae of statistical procedures.

Spearman's reification of g

Spearman could not rest content with the idea that he had probed deeply under the empirical results of mental tests and

found a single abstract factor underlying all performance. Nor could he achieve adequate satisfaction by identifying that factor with what we call intelligence itself.* Spearman felt compelled to ask more of his *g*: it must measure some physical property of the brain; it must be a "thing" in the most direct, material sense. Even if neurology had found no substance to identify with *g*, the brain's performance on mental tests proved that such a physical substrate must exist. Thus, caught up in physics envy again, Spearman described his own "adventurous step of deserting all actually observable phenomena of the mind and proceeding instead to invent an underlying something which—by analogy with physics—has been called mental energy" (1927, p. 89).

Spearman looked to the basic property of *g*—its influence in varying degree, upon mental operations—and tried to imagine what physical entity best fitted such behavior. What else, he argued, but a form of energy pervading the entire brain and activating a set of specific "engines," each with a definite locus. The more energy, the more general activation, the more intelligence. Spearman wrote (1923, p. 5):

This continued tendency to success of the same person throughout all variations of both form and subject matter—that is to say, throughout all conscious aspects of cognition whatever—appears only explicable by some factor lying deeper than the phenomena of consciousness. And thus there emerges the concept of a hypothetical general and purely quantitative factor underlying all cognitive performances of any kind. . . . The factor was taken, pending further information, to consist in something of the nature of an "energy" or "power" which serves in common the whole cortex (or possibly, even, the whole nervous system)."

If *g* pervades the entire cortex as a general energy, then the *s*-factors for each test must have more definite locations. They must represent specific groups of neurons, activated in different ways by the energy identified with *g*. The *s*-factors, Spearman wrote (and not merely in metaphor), are engines fueled by a circulating *g*.

Each different operation must necessarily be further served by some specific factor peculiar to it. For this factor also, a physiological substrate has been suggested, namely the particular group of neurons specially serv-

*At least in his early work. Later, as we have seen, he abandoned the word intelligence as a result of its maddening ambiguity in common usage. But he did not cease to regard *g* as the single cognitive essence that should be called intelligence, had not vernacular (and technical) confusion made such a mockery of the term.

ing the particular kind of operation. These neural groups would thus function as alternative "engines" into which the common supply of "energy" could be alternatively distributed. Successful action would always depend, partly on the potential of energy developed in the whole cortex, and partly on the efficiency of the specific group of neurons involved. The relative influence of these two factors could vary greatly according to the kind of operation; some kinds would depend more on the potential of the energy, others more on the efficiency of the engine (1923, pp. 5-6).

The differing *g*-loadings of tests had been provisionally explained: one mental operation might depend primarily upon the character of its engine (high *s* and low *g*-loading), another might owe its status to the amount of general energy involved in activating its engine (high *g*-loading).

Spearman felt sure that he had discovered the basis of intelligence, so sure that he proclaimed his concept impervious to disproof. He expected that a physical energy corresponding with *g* would be found by physiologists: "There seem to be grounds for hoping that a material energy of the kind required by psychologists will some day actually be discovered" (1927, p. 407). In this discovery, Spearman proclaimed, "physiology will achieve the greatest of its triumphs" (1927, p. 408). But should no physical energy be found, still an energy there must be—but of a different sort:

And should the worst arrive and the required physiological explanation remain to the end undiscoverable, the mental facts will none the less remain facts still. If they are such as to be best explained by the concept of an underlying energy, then this concept will have to undergo that which after all is only what has long been demanded by many of the best psychologists—it will have to be regarded as purely mental (1927, p. 408).

Spearman, in 1927 at least, never considered the obvious alternative: that his attempt to reify *g* might be invalid in the first place.

Throughout his career, Spearman tried to find other regularities of mental functioning that would validate his theory of general energy and specific engines. He enunciated (1927, p. 133) a "law of constant output" proclaiming that the cessation of any mental activity causes others of equal intensity to commence. Thus, he reasoned, general energy remains intact and must always be activating something. He found, on the other hand, that fatigue is "selectively transferred"—that is, tiring in one mental activity entails fatigue in some related areas, but not in others (1927, p. 318). Thus, fatigue

cannot be attributed to "decrease in the supply of the general psycho-physiological energy," but must represent a build up of toxins that act selectively upon certain kinds of neurons. Fatigue, Spearman proclaimed, "primarily concerns not the energy but the engines" (1927, p. 318).

Yet, as we find so often in the history of mental testing, Spearman's doubts began to grow until he finally recanted in his last (posthumously published) book of 1950. He seemed to pass off the theory of energy and engines as a folly of youth (though he had defended it staunchly in middle age). He even abandoned the attempt to reify factors, recognizing belatedly that a mathematical abstraction need not correspond with a physical reality. The great theorist had entered the camp of his enemies and recast himself as a cautious empiricist (1950, p. 25):

We are under no obligation to answer such questions as: whether "factors" have any "real" existence? do they admit of genuine "measurement"? does the notion of "ability" involve at bottom any kind of cause, or power? Or is it only intended for the purpose of bare description? . . . At their time and in their place such themes are doubtless well enough. The senior writer himself has indulged in them not a little. *Dulce est desipere in loco* [it is pleasant to act foolishly from time to time—a line from Horace]. But for the present purposes he has felt himself constrained to keep within the limits of barest empirical science. These he takes to be at bottom nothing but description and prediction. . . . The rest is mostly illumination by way of metaphor and similes.

The history of factor analysis is strewn with the wreckage of misguided attempts at reification. I do not deny that patterns of causality may have identifiable and underlying, physical reasons, and I do agree with Eysenck when he states (1953, p. 113): "Under certain circumstances, factors may be regarded as hypothetical causal influences underlying and determining the observed relationships between a set of variables. It is only when regarded in this light that they have interest and significance for psychology." My complaint lies with the practice of assuming that the mere existence of a factor, in itself, provides a license for causal speculation. Factorists have consistently warned against such an assumption, but our Platonic urges to discover underlying essences continue to prevail over proper caution. We can chuckle, with the beneficence of hindsight, at psychiatrist T. V. Moore who, in 1933, postulated def-

inite genes for catatonic, deluded, manic, cognitive, and constitutional depression because his factor analysis grouped the supposed measures of these syndromes on separate axes (in Wolfle, 1940). Yet in 1972 two authors found an association of dairy production with florid vocalization on the tiny thirteenth axis of a nineteen-axis factor analysis for musical habits of various cultures—and then suggested "that this extra source of protein accounts for many cases of energetic vocalizing" (Lomax and Berkowitz, 1972, p. 232).

Automatic reification is invalid for two major reasons. First, as I discussed briefly on pp. 282–285 and will treat in full on pp. 326–347, no set of factors has any claim to exclusive concordance with the real world. Any matrix of positive correlation coefficients can be factored, as Spearman did, into g and a set of subsidiary factors or, as Thurstone did, into a set of "simple structure" factors that usually lack a single dominant direction. Since either solution resolves the same amount of information, they are equivalent in mathematical terms. Yet they lead to contrary psychological interpretations. How can we claim that one, or either, is a mirror of reality?

Second, any single set of factors can be interpreted in a variety of ways. Spearman read his strong g as evidence for a single reality underlying all cognitive mental activity, a general energy within the brain. Yet Spearman's most celebrated English colleague in factor analysis, Sir Godfrey Thomson, accepted Spearman's mathematical results but consistently chose to interpret them in an opposite manner. Spearman argued that the brain could be divided into a set of specific engines, fueled by a general energy. Thomson, using the same data, inferred that the brain has hardly any specialized structure at all. Nerve cells, he argued, either fire completely or not at all—they are either off or on, with no intermediary state. Every mental test samples a random array of neurons. Tests with high g -loadings catch many neurons in the active state; others, with low g -loadings, have simply sampled a smaller amount of unstructured brain. Thomson concluded (1939): "Far from being divided up into a few 'unitary factors,' the mind is a rich, comparatively undifferentiated complex of innumerable influences—on the physiological side an intricate network of possibilities of intercommunication." If the same mathematical pattern can yield such disparate interpretations, what claim can either have upon reality?

Spearman on the inheritance of g

Two of Spearman's primary claims appear in most hereditarian theories of mental testing: the identification of intelligence as a unitary "thing," and the inference of a physical substrate for it. But these claims do not complete the argument: a single, physical substance may achieve its variable strength through effects of environment and education, not from inborn differences. A more direct argument for the heritability of *g* must be made, and Spearman supplied it.

The identification of *g* and *s* with energy and engines again provided Spearman with his framework. He argued that the *s*-factors record training in education, but that the strength of a person's *g* reflects heredity alone. How can *g* be influenced by education, Spearman argued (1927, p. 392), if *g* ceases to increase by about age sixteen but education may continue indefinitely thereafter? How can *g* be altered by schooling if it measures what Spearman called *eduction* (or the ability to synthesize and draw connections) and not *retention* (the ability to learn facts and remember them)—when schools are in the business of imparting information? The engines can be stuffed full of information and shaped by training, but the brain's general energy is a consequence of its inborn structure:

The effect of training is confined to the specific factor and does not touch the general one; physiologically speaking, certain neurons become habituated to particular kinds of action, but the free energy of the brain remains unaffected. . . . Though unquestionably the development of specific abilities is in large measure dependent upon environmental influences, that of general ability is almost wholly governed by heredity (1914, pp. 233–234).

IQ, as a measure of *g*, records an innate general intelligence; the marriage of the two great traditions in mental measurement (IQ testing and factor analysis) was consummated with the issue of heredity.

On the vexatious issue of group differences, Spearman's views accorded with the usual beliefs of leading western European male scientists at the time (see Fig. 6.9). Of blacks, he wrote (1927, p. 379), invoking *g* to interpret the army mental tests:

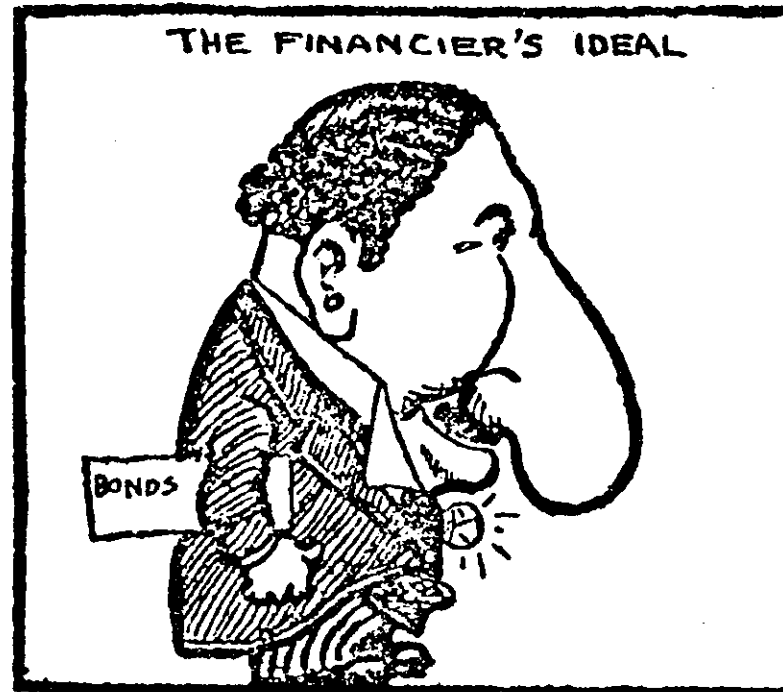
On the average of all the tests, the colored were about two years behind the white; their inferiority extended through all ten tests, but it was most marked in just those which are known to be most saturated with *g*.

In other words, blacks performed most poorly on tests having strongest correlations with *g*, or innate general intelligence.

Of whites from southern and eastern Europe, Spearman wrote (1927, p. 379), praising the American Immigration Restriction Act of 1924:

The general conclusion emphasized by nearly every investigator is that, as regards "intelligence," the Germanic stock has on the average a marked advantage over the South European. And this result would seem to have

6•9 Racist stereotype of a Jewish financier, reproduced from the first page of Spearman's 1914 article (see Bibliography). Spearman used this figure to criticize beliefs in group factors for such particular items of intellect, but its publication illustrates the acceptable attitudes of another age.



had vitally important practical consequences in shaping the recent very stringent American laws as to admission of immigrants.

Yet it would be incorrect to brand Spearman as an architect of the hereditarian theory for differences in intelligence among human groups. He supplied some important components, particularly the argument that intelligence is an innate, single, scorable "thing." He also held conventional views on the source of average differences in intelligence between races and national groups. But he did not stress the ineluctability of differences. In fact, he attributed sexual differences to training and social convention (1927, p. 229) and had rather little to say about social classes. Moreover, when discussing racial differences, he always coupled his hereditarian claim about average scores with an argument that the range of variation within any racial or national group greatly exceeds the small average difference between groups—so that many members of an "inferior" race will surpass the average intelligence of a "superior" group (1927, p. 380, for example).*

Spearman also recognized the political force of hereditarian claims, though he did not abjure either the claim or the politics: "All great efforts to improve human beings by way of training are thwarted through the apathy of those who hold the sole feasible road to be that of stricter breeding" (1927, p. 376).

But, most importantly, Spearman simply didn't seem to take much interest in the subject of hereditary differences among peoples. While the issue swirled about him and buried his profession in printer's ink, and while he himself had supplied a basic argument for the hereditarian school, the inventor of *g* stood aside in apparent apathy. He had studied factor analysis because he wanted to understand the structure of the human brain, not as a guide to measuring differences between groups, or even among individuals. Spearman may have been a reluctant courtier, but the politically potent union of IQ and factor analysis into a hereditarian theory of intelligence was engineered by Spearman's successor in the chair of psychology at University College—Cyril Burt. Spearman may have cared little, but the innate character of intelligence was the *idée fixe* of Sir Cyril's life.

* Richard Herrnstein and Charles Murray emphasize the same arguments to obviate a charge of racism against *The Bell Curve* (1994)—see first two essays at end of book.

Cyril Burt and the hereditarian synthesis

The source of Burt's uncompromising hereditarianism

Cyril Burt published his first paper in 1909. In it, he argued that intelligence is innate and that differences between social classes are largely products of heredity; he also cited Spearman's *g* as primary support. Burt's last paper in a major journal appeared posthumously in 1972. It sang the very same tune: intelligence is innate and the existence of Spearman's *g* proves it. For all his more dubious qualities, Cyril Burt certainly had staying power. The 1972 paper proclaims:

The two main conclusions we have reached seem clear and beyond all question. The hypothesis of a general factor entering into every type of cognitive process, tentatively suggested by speculations derived from neurology and biology, is fully borne out by the statistical evidence; and the contention that differences in this general factor depend largely on the individual's genetic constitution appears incontestable. The concept of an innate, general, cognitive ability, which follows from these two assumptions, though admittedly a sheer abstraction, is thus wholly consistent with the empirical facts (1972, p. 188).

Only the intensity of Sir Cyril's adjectives had changed. In 1912 he had termed this argument "conclusive"; by 1972 it had become "incontestable."

Factor analysis lay at the core of Burt's definition of intelligence as i.g.c. (innate, general, cognitive) ability. In his major work on factor analysis (1940, p. 216), Burt developed his characteristic use of Spearman's thesis. Factor analysis shows that "a *general* factor enters into all *cognitive* processes," and "this general factor appears to be largely, if not wholly, inherited or *innate*"—again, i.g.c. ability. Three years earlier (1937, pp. 10–11) he had tied *g* to an ineluctable heredity even more graphically:

This general intellectual factor, central and all-pervading, shows a further characteristic, also disclosed by testing and statistics. It appears to be inherited, or at least inborn. Neither knowledge nor practice, neither interest nor industry, will avail to increase it.

Others, including Spearman himself, had drawn the link between *g* and heredity. Yet no one but Sir Cyril ever pursued it with such stubborn, almost obsessive gusto: and no one else

wielded it as such an effective political tool. The combination of hereditarian bias with a reification of intelligence as a single, measurable entity defined Burt's unyielding position.

I have discussed the roots of the second component: intelligence as a reified factor. But where did the first component—rigid hereditarianism—arise in Burt's view of life? It did not flow logically from factor analysis itself, for it cannot (see pp. 280–282). I will not attempt to answer this question by referring either to Burt's psyche or his times (though Hearnshaw, 1979, has made some suggestions). But I will demonstrate that Burt's hereditarian argument had no foundation in his empirical work (either honest or fraudulent), and that it represented an a priori bias imposed upon the studies that supposedly proved it. It also acted, through Burt's zealous pursuit of his *idée fixe*, as a distorter of judgment and finally as an incitement to fraud.*

BURT'S INITIAL "PROOF" OF INNATENESS

Throughout his long career, Burt continually cited his first paper of 1909 as a proof that intelligence is innate. Yet the study falters both on a flaw of logic (circular reasoning) and on the remarkably scant and superficial character of the data themselves. This publication proves only one thing about intelligence—that Burt began his study with an a priori conviction of its innateness, and reasoned back in a vicious circle to his initial belief. The "evidence"—what there was of it—served only as selective window dressing.

At the outset of his 1909 paper, Burt set three goals for himself. The first two reflect the influence of Spearman's pioneering work in factor analysis ("can general intelligence be detected and measured"; "can its nature be isolated and its meaning analyzed"). The third represents Burt's peculiar concern: "Is its development predominantly determined by environmental influence and individual acquisition, or is it rather dependent upon the inheritance of a racial character or family trait" (1909, p. 96).

Not only does Burt proclaim this third question "in many ways

* Of Burt's belief in the innateness of intelligence, Hearnshaw writes (1979, p. 49): "It was for him almost an article of faith, which he was prepared to defend against all opposition, rather than a tentative hypothesis to be refuted, if possible, by empirical tests. It is hard not to feel that almost from the first Burt showed an excessive assurance in the finality and correctness of his conclusions."

the most important of all," but he also gives away his answer in stating why we should be so concerned. Its importance rests upon:

... the growing belief that innate characters of the family are more potent in evolution than the acquired characters of the individual, the gradual apprehension that unsupplemented humanitarianism and philanthropy may be suspending the natural elimination of the unfit stocks—these features of contemporary sociology make the question whether ability is inherited one of fundamental moment (1909, p. 169).

Burt selected forty-three boys from two Oxford schools, thirty sons of small tradesmen from an elementary school and thirteen upper-class boys from preparatory school. In this "experimental demonstration that intelligence is hereditary" (1909, p. 179), with its ludicrously small sample, Burt administered twelve tests of "mental functions of varying degrees of complexity" to each boy. (Most of these tests were not directly cognitive in the usual sense, but more like the older Galtonian tests of physiology—attention, memory, sensory discrimination, and reaction time). Burt then obtained "careful empirical estimates of intelligence" for each boy. This he did not by rigorous Binet testing, but by asking "expert" observers to rank the boys in order of their intelligence independent of mere school learning. He obtained these rankings from the headmasters of the schools, from teachers, and from "two competent and impartial boys" included in the study. Writing in the triumphant days of British colonialism and derring-do, Burt instructed his two boys on the meaning of intelligence:

Supposing you had to choose a leader for an expedition into an unknown country, which of these 30 boys would you select as the most intelligent? Failing him, which next? (1909, p. 106)

Burt then searched for correlations between performance on the twelve tests and the rankings produced by his expert witnesses. He found that five tests had correlation coefficients with intelligence above 0.5, and that poorest correlations involved tests of "lower senses—touch and weight," while the best correlations included tests of clearer cognitive import. Convinced that the twelve tests measured intelligence, Burt then considered the scores themselves. He found that the upper-class boys performed better than the lower-middle-class boys in all tests save those involving weight and touch. The upper-class boys must therefore be smarter.

But is the superior smartness of upper-class boys innate or

acquired as a function of advantages in home and schooling? Burt gave four arguments for discounting environment:

1. The environment of lower-middle-class boys cannot be poor enough to make a difference since their parents can afford the ninepence a week required to attend school: "Now in the case of the lowest social classes, general inferiority at mental tests might be attributable to unfortunate environmental and post-natal influences. . . . But such conditions could not be suspected with the boys who, at a fee of 9d a week, attended the Central Elementary School" (1909, p. 173). In other words, environment can't make a difference until it reduces a child to near starvation.

2. The "educative influences of home and social life" seem small. In making this admittedly subjective assessment, Burt appealed to a fine intuition honed by years of gut-level experience. "Here, however, one must confess, such speculative arguments can convey little conviction to those who have not witnessed the actual manner of the respective boys."

3. The character of the tests themselves precludes much environmental influence. As tests of sensation and motor performance, they do not involve "an appreciable degree of acquired skill or knowledge. . . . There is reason, therefore, to believe that the differences revealed are mainly innate" (1909, p. 180).

4. A retesting of the boys eighteen months later, after several had entered professions or new schools, produced no important readjustment of ranks. (Did it ever occur to Burt that environment might have its primary influence in early life, and not only in immediate situations?)

The problem with all these points, and with the design of the entire study, is a patent circularity in argument. Burt's claim rested upon correlations between test performances and a ranking of intelligence compiled by "impartial" observers. (Arguments about the "character" of the tests themselves are secondary, for they would count for nothing in Burt's design if the tests did not correlate with independent assessments of intelligence.) We must know what the subjective rankings mean in order to interpret the correlations and make any use of the tests themselves. For if the rankings of teachers, headmasters, and colleagues, however sincerely attempted, record the advantages of upbringing more than the differential blessings of genetics, then the ranks are primarily a record

of environment, and the test scores may provide just another (and more imperfect) measure of the same thing. Burt used the correlation between two criteria as evidence for heredity without ever establishing that either criterion measured his favored property.

In any case, all these arguments for heredity are indirect. Burt also claimed, as his final proof, a direct test of inheritance: the boys' measured intelligence correlated with that of their parents:

Wherever a process is correlated with intelligence, these children of superior parentage resemble their parents in being themselves superior. . . . Proficiency at such tests does not depend upon opportunity or training, but upon some quality innate. The resemblance in degree of intelligence between the boys and their parents must, therefore, be due to inheritance. We thus have an experimental demonstration that intelligence is hereditary (1909, p. 181).

But how did Burt measure parental intelligence? The answer, remarkable even from Burt's point of view, is that he didn't: he merely assumed it from profession and social standing. Intellectual, upper-class parents must be innately smarter than tradesmen. But the study was designed to assess whether or not performance on tests reflects inborn qualities or the advantages of social standing. One cannot, therefore, turn around and infer intelligence directly from social standing.

We know that Burt's later studies of inheritance were fraudulent. Yet his early and honest work is riddled with flaws so fundamental that they stand in scarcely better light. As in the 1909 study, Burt continually argued for innateness by citing correlations in intelligence between parents and offspring. And he continually assessed parental intelligence by social standing, not by actual tests.

For example, after completing the Oxford study, Burt began a more extensive program of testing in Liverpool. He cited high correlations between parents and offspring as a major argument for innate intelligence, but never provided parental scores. Fifty years later, L. S. Penrose read Burt's old work, noted the absent data, and asked Burt how he had measured parental intelligence. The old man replied (in Hearnshaw, 1979, p. 29):

The intelligence of the parents was assessed primarily on the basis of their actual jobs, checked by personal interviews; about a fifth were also tested to standardize the impressionistic assessments.

Hearnshaw comments (1979, p. 30): "Inadequate reporting and incautious conclusions mark this first incursion of Burt into the genetic field. We have here, right at the beginning of his career, the seeds of later troubles."

Even when Burt did test subjects, he rarely reported the actual scores as measured, but "adjusted" them according to his own assessment of their failure to measure true intelligence as he and other experts subjectively judged it. He admitted in a major work (1921, p. 280):

I did not take my test results just as they stood. They were carefully discussed with teachers, and freely corrected whenever it seemed likely that the teacher's view of the relative merits of his own pupils gave a better estimate than the crude test marks.

Such a procedure is not without its commendable intent. It does admit the inability of a mere number, calculated during a short series of tests, to capture such a subtle notion as intelligence. It does grant to teachers and others with extensive personal knowledge the opportunity to record their good judgment. But it surely makes a mockery of any claim that a specific hypothesis is under objective and rigorous test. For if one believes beforehand that well-bred children are innately intelligent, then in what direction will the scores be adjusted?*

Despite his minuscule sample, his illogical arguments, and his dubious procedures, Burt closed his 1909 paper with a statement of personal triumph (p. 176):

Parental intelligence, therefore, may be inherited, individual intelligence measured, and general intelligence analyzed; and they can be analyzed, measured and inherited to a degree which few psychologists have hitherto legitimately ventured to maintain.

When Burt recycled these data in a 1912 paper for the *Eugenics Review*, he added additional "proof" with even smaller samples. He

*Sometimes, Burt descended even further into circular illogic and claimed that tests must measure innate intelligence because the testers constructed them to do so: "Indeed from Binet onwards practically all the investigators who have attempted to construct 'intelligence tests' have been primarily searching for some measure of *inborn* capacity, as distinct from acquired knowledge or skill. With such an interpretation it obviously becomes foolish to inquire how far 'intelligence' is due to environment and how far it is due to innate constitution: the very definition begs and settles the question" (1943, p. 88).

discussed Alfred Binet's two daughters, noted that their father had been disinclined to connect physical signs with mental prowess, and pointed out that the blond, blue-eyed, large-headed daughter of Teutonic appearance was objective and forthright, while the darker daughter tended to be impractical and sentimental. *Touché*.

Burt was no fool. I confess that I began reading him with the impression, nurtured by spectacular press reports of his fraudulent work, that he was simply a devious and foxy charlatan. To be sure, that he became and for complex reasons (see pp. 264-269). But as I read, I gained respect for Burt's enormous erudition, for his remarkable sensitivity in most areas, and for the subtlety and complexity of his reasoning; I ended up liking most things about him in spite of myself. And yet, this assessment makes the extraordinary weakness of his reasoning about the innateness of intelligence all the more puzzling. If he had simply been a fool, then foolish arguments would denote consistency of character.

My dictionary defines an *idée fixe*, or fixed idea, as "a persistent or obsessing idea, often delusional, from which a person cannot escape." The innateness of intelligence was Burt's *idée fixe*. When he turned his intellectual skills to other areas, he reasoned well, subtly, and often with great insight. When he considered the innateness of intelligence, blinders descended and his rational thinking evaporated before the hereditarian dogma that won his fame and eventually sealed his intellectual doom. It may be remarkable that Burt could operate with such a duality in styles of reasoning. But I find it much more remarkable that so many others believed Burt's statements about intelligence when his arguments and data, all readily available in popular publications, contained such patent errors and specious claims. What does this teach us about shared dogma masquerading as objectivity?

LATER ARGUMENTS

Perhaps I have been unfair in choosing Burt's earliest work for criticism. Perhaps the foolishness of youth soon yielded to mature wisdom and caution. Not at all; Burt was nothing if not ontogenetically consistent. The argument of 1909 never changed, never gained subtlety, and ended with manufactured support. The innateness of intelligence continued to function as dogma. Consider the primary argument of Burt's most famous book, *The Back-*

ward Child (1937), written at the height of his powers and before his descent into conscious fraud.

Backwardness, Burt notes, is defined by achievement in school, not by tests of intelligence: backward children are more than a year behind in their schoolwork. Burt argues that environmental effects, if at all important, should have most impact upon children in this category (those much further behind in school are more clearly genetically impaired). Burt therefore undertook a statistical study of environment by correlating the percentage of backward children with measures of poverty in the boroughs of London. He calculated an impressive array of strong correlations: 0.73 with percentage of people below the poverty line, 0.89 with overcrowding, 0.68 with unemployment, and 0.93 with juvenile mortality. These data seem to provide a prima-facie case for a dominant environmental influence upon backwardness, but Burt demurs. There is another possibility. Perhaps the innately poorest stocks create and then gravitate to the worst boroughs, and degree of poverty is merely an imperfect measure of genetic worthlessness.

Burt, guided by his *idée fixe*, opted for innate stupidity as the primary cause of poverty (1937, p. 105). He invoked IQ testing as his major argument. Most backward children score 1 to 2 standard deviations below the mean (70–85), within a range technically designated as "dull." Since IQ records innate intelligence, most backward children perform poorly in school because they are dull, not (or only indirectly) because they are poor. Again, Burt rides his circle. He wishes to prove that deficiency of innate intelligence is the major cause of poor performance in school. He knows full well that the link between IQ score and innateness is an unresolved issue in intense debates about the meaning of IQ—and he admits in many places that the Stanford-Binet test is, at best, only an imperfect measure of innateness (e.g., 1921, p. 90). Yet, using the test scores as a guide, he concludes:

In well over half the cases, the backwardness seems due chiefly to intrinsic mental factors; here, therefore, it is primary, innate, and to that extent beyond all hope of cure (1937, p. 110).

Consider Burt's curious definition of innate in this statement. An innate character, as inborn and, in Burt's usage, inherited, forms part of an organism's biological constitution. But the demonstra-

tion that a trait represents nature unaffected by nurture does not guarantee its ineluctable state. Burt inherited poor vision. No doctor ever rebuilt his eyes to an engineer's paradigm of normal design, but Burt wore eyeglasses and the only clouding of his vision was conceptual.

The Backward Child also abounds in tangential statements that record Burt's hereditarian biases. He writes about an environmental handicap—recurrent catarrh among the poor—and discusses hereditary susceptibility (quite plausible) with an arresting quip for graphic emphasis:

... exceptionally prevalent in those whose faces are marked by developmental defects—by the round receding forehead, the protruding muzzle, the short and upturned nose, the thickened lips, which combine to give to the slum child's profile a negroid or almost simian outline. . . . "Apes that are hardly anthropoid" was the comment of one headmaster, who liked to sum up his cases in a phrase (1937, p. 186).

He wonders about the intellectual achievement of Jews and attributes it, in part, to inherited myopia that keeps them off the playing fields and adapts them for poring over account books.

Before the invention of spectacles, the Jew whose living depended upon his ability to keep accounts and read them, would have been incapacitated by the age of 50, had he possessed the usual tendency to hypermetropia: on the other hand (as I can personally testify) the myope . . . can dispense with glasses for near work without much loss of efficiency (1937, p. 219).

BURT'S BLINDNESS

The blinding power of Burt's hereditarian biases can best be appreciated by studying his approach to subjects other than intelligence. For here he consistently showed a commendable caution. He recognized the complexity of causation and the subtle influence that environment can exert. He railed against simplistic assumptions and withheld judgment pending further evidence. Yet as soon as Burt returned to his favorite subject of intelligence, the blinders descended and the hereditarian catechism came forward again.

Burt wrote with power and sensitivity about the debilitating effects of poor environments. He noted that 23 percent of the cockney youth he interviewed had never seen a field or a patch of

grass, not "even in a Council park," 64 percent had never seen a train, and 98 percent had never seen the sea. The following passage displays a measure of paternalistic condescension and stereotyping, but it also presents a powerful image of poverty in working-class homes, and its intellectual effect upon children (1937, p. 127).

His mother and father know astonishingly little of any life except their own, and have neither the time nor the leisure, neither the ability nor the disposition, to impart what little they know. The mother's conversation may be chiefly limited to the topics of cleaning, cooking, and scolding. The father, when not at work, may spend most of his time "round the corner" refreshing a worn-out body, or sitting by the fire with cap on and coat off, sucking his pipe in gloomy silence. The vocabulary that the child absorbs is restricted to a few hundred words, most of them inaccurate, uncouth, or mispronounced, and the rest unfit for reproduction in the schoolroom. In the home itself there is no literature that deserves the title; and the child's whole universe is closed in and circumscribed by walls of brick and a pall of smoke. From one end of the year to the other, he may go no farther than the nearest shops or the neighborhood recreation ground. The country or the seaside are mere words to him, dimly suggesting some place to which cripples are sent after an accident, visualized perhaps in terms of some photographic "souvenir from Southend" or some pictorial "memento from Margate," all framed in shells, brought back by his parents on a bank-holiday trip a few weeks after their wedding.

Burt appended this comment from a "burly bus conductor" to his description: "Book learning isn't for kids that'll have to earn their bread. It's only for them as likes to give themselves the hairs of the 'ighbrow."

Burt could apply what he understood so well to subjects other than intelligence. Consider his views on juvenile delinquency and left-handedness. Burt wrote extensively on the cause of delinquency and attributed it to complex interactions between children and their environment: "The problem never lies in the 'problem child' alone: it lies always in the relations between that child and his environment" (1940, p. 243). If poor behavioral performance merits such an assessment, why not say the same about poor intellectual performance? One might suspect that Burt relied again upon test scores, arguing that delinquents tested well and could not be misbehaving as a result of innate stupidity. But, in fact, delinquents often tested as badly as poor children regarded by Burt as innately deficient in intelligence. Yet Burt recognized that IQ scores of

delinquents may not reflect inherited ability because they rebel against taking the tests:

For what to them must seem nothing but a resuscitated school examination, delinquents, as a rule, feel little inclination and much distaste. From the outset they assume they are more likely to fail than succeed, more likely to be reproached than commended. . . . Unless, indeed, to circumvent their suspicion and secure their good-will special manoeuvres be tactfully tried, their apparent prowess with all such tests will fall much below their veritable powers. . . . In the causation of juvenile delinquency . . . the share contributed by mental defect has unquestionably been magnified by those who, trusting so exclusively to the Binet-Simon scale, have ignored the factors which depreciate its results (1921, pp. 189-190).

But why not say that poverty often entails a similar disinclination and sense of defeat?

Burt (1937, p. 270) regarded left-handedness as the "motor disability . . . which interferes most widely with the ordinary tasks of the classroom." As chief psychologist of the London schools, he therefore devoted much study to its cause. Unburdened by a priori conviction in this case, he devised and attempted to test a wide range of potential environmental influences. He studied medieval and Renaissance paintings to determine if Mary usually carried the infant Jesus on her right hip. If so, babies would wrap their left arms about their mother's neck, leaving their right hand free for more dextrous (literally right-handed) motion. He wondered if greater frequency of right-handedness might record the asymmetry of internal organs and the need for protection imposed by our habits. If heart and stomach lie to the left of the midline, then a warrior or worker would naturally turn his left side away from potential danger, "trust to the more solid support of the right side of the trunk, and so use his right hand and arm for wielding heavy instruments and weapons" (1937, p. 270). In the end, Burt opted for caution and concluded that he could not tell:

I should in the last resort contend that probably all forms of left-handedness are only indirectly hereditary: postnatal influence seems always to enter in. . . . I must accordingly repeat that, here as elsewhere in psychology, our present knowledge is far too meager to allow us to declare with any assurance what is inborn and what is not (1937, pp. 303-304).

Substitute "intelligence" for "left-handedness" and the statement is a model of judicious inference. In fact, left-handedness is more

clearly an entity than intelligence, and probably more subject to definite and specifiable hereditary influence. Yet here, where his case for innateness was better, Burt tested all the environmental influences—some rather farfetched—that he could devise, and finally declared the subject too complex for resolution.

BURT'S POLITICAL USE OF INNATENESS

Burt extended his belief in the innateness of individual intelligence to only one aspect of average differences between groups. He did not feel (1912) that races varied much in inherited intelligence, and he argued (1921, p. 197) that the different behaviors of boys and girls can be traced largely to parental treatment. But differences in social class, the wit of the successful and dullness of the poor, are reflections of inherited ability. If race is America's primary social problem, then class has been Britain's corresponding concern.

In his watershed* paper (1943) on "ability and income," Burt concludes that "the wide inequality in personal income is largely, though not entirely, an indirect effect of the wide inequality in innate intelligence." The data "do not support the view (still held by many educational and social reformers) that the apparent inequality in intelligence of children and adults is in the main an indirect consequence of inequality in economic conditions" (1943, p. 141).

Burt often denied that he wished to limit opportunities for achievement by regarding tests as measures of innate intelligence. He argued, on the contrary, that tests could identify those few individuals in the lower classes whose high innate intelligence would not otherwise be recognized under a veneer of environmental disadvantage. For "among nations, success in the struggle for survival is bound to depend more and more on the achievements of a small handful of individuals who are endowed by nature with outstanding gifts of ability and character" (1959, p. 31). These people must be identified and nurtured to compensate for "the comparative ineptitude of the general public" (1959, p. 31). They must be encouraged and rewarded, for the rise and fall of a nation does not depend upon genes peculiar to an entire race, but upon

*Hearnshaw (1979) suspects that this paper marks Burt's first use of fraudulent data.

"changes in the relative fertility of its leading members or its leading classes" (1962, p. 49).

Tests may have been the vehicle by which a few children escaped from the strictures of a fairly inflexible class structure. But what was their effect on the vast majority of lower-class children whom Burt unfairly branded as unable, by inheritance, ever to develop much intelligence—and therefore undeserving, by reason, of higher social standing?

Any recent attempt to base our educational policy for the future on the assumption that there are no real differences, or at any rate no important differences, between the average intelligence of the different social classes, is not only bound to fail; it is likely to be fraught with disastrous consequences for the welfare of the nation as a whole, and at the same time to result in needless disappointments for the pupils concerned. The facts of genetic inequality, whether or not they conform to our personal wishes and ideals, are something that we cannot escape (1959, p. 28). . . . A definite limit to what children can achieve is inexorably set by the limitations of their innate capacity (1969).

Burt's extension of Spearman's theory

Cyril Burt may be known best to the public as a hereditarian in the field of mental testing, but his reputation as a theoretical psychologist rested primarily upon his work in factor analysis. He did not invent the technique, as he later claimed; but he was Spearman's successor, both literally and figuratively, and he became the leading British factorist of his generation.

Burt's genuine achievements in factor analysis were substantial. His complex and densely reasoned book on the subject (1940) was the crowning achievement of Spearman's school. Burt wrote that it "may prove to be a more lasting contribution to psychology than anything else I have yet written" (letter to his sister quoted in Hearnshaw, 1979, p. 154). Burt also pioneered (though he did not invent) two important extensions of Spearman's approach—an inverted technique (discussed on pp. 322–323) that Burt called "correlation between persons" (now known to aficionados as "Q-mode factor analysis"), and an expansion of Spearman's two-factor theory to add "group factors" at a level between *g* and *s*.

Burt toed Spearman's line in his first paper of 1909. Spearman had insisted that each test recorded only two properties of mind—

a general factor common to all tests and a specific factor peculiar to that test alone. He denied that clusters of tests showed any significant tendency to form "group factors" between his two levels—that is, he found no evidence for the "faculties" of an older psychology, no clusters representing verbal, spatial, or arithmetic ability, for example. In his 1909 paper, Burt did note a "discernible, but small" tendency for grouping in allied tests. But he proclaimed it weak enough to ignore ("vanishingly minute" in his words), and argued that his results "confirm and extend" Spearman's theory.

But Burt, unlike Spearman, was a practitioner of testing (responsible for all of London's schools). Further studies in factor analysis continued to distinguish group factors, though they were always subsidiary to *g*. As a practical matter for guidance of pupils, Burt realized that he could not ignore the group factors. With a purely Spearmanian approach, what could a pupil be told except that he was generally smart or dumb? Pupils had to be guided toward professions by identifying strengths and weaknesses in more specific areas.

By the time Burt did his major work in factor analysis, Spearman's cumbersome method of tetrad differences had been replaced by the principal components approach outlined on pp. 275–280. Burt identified group factors by studying the projection of individual tests upon the second and subsequent principal components. Consider Fig. 6.6: In a matrix of positive correlation coefficients, vectors representing individual tests are all clustered together. The first principal component, Spearman's *g* runs through the middle of the cluster and resolves more information than any other axis could. Burt recognized that no consistent patterns would be found on subsequent axes if Spearman's two-factor theory held—for the vectors would not form subclusters if their only common variance had already been accounted for by *g*. But if the vectors form subclusters representing more specialized abilities, then the first principal component must run *between* the subclusters if it is to be the best average fit to all vectors. Since the second principal component is perpendicular to the first, some subclusters must project positively upon it and others negatively (as Fig. 6.6 shows with its negative projections for verbal tests and positive projections for arithmetic tests). Burt called these axes *bipolar factors*, because they included clusters of positive and negative pro-

jections. He identified as *group factors* the clusters of positive and negative projections themselves.

Burt's identification of group factors may seem, superficially, to challenge Spearman's theory, but in fact it provided an extension and improvement that Spearman eventually welcomed. The essence of Spearman's claim is the primacy of *g*, and the subordination of all other determinants of intelligence to it. Burt's identification of group factors preserved this notion of hierarchy, and extended it by adding another level between *g* and *s*. In fact, Burt's treatment of group factors as a level in a hierarchy subordinate to *g* saved Spearman's theory from the data that seemed to threaten it. Spearman originally denied group factors, but evidence for them continued to accumulate. Many factorists began to view this evidence as a denigration of *g* and as a wedge for toppling Spearman's entire edifice. Burt strengthened the building, preserved the preeminent role of *g*, and extended Spearman's theory by enumerating further levels subordinate to *g*. The factors, Burt wrote (1949, p. 199), are "organized on what may be called a hierarchical basis. . . . There is first a comprehensive general factor, covering all cognitive activities; next a comparatively small number of broad group factors, covering different abilities classified according to their form or content. . . . The whole series appears to be arranged on successive levels, the factors on the lowest level being the most specific and the most numerous of all."

Spearman had advocated a two-factor theory; Burt proclaimed a four-factor theory: the *general* factor or Spearman's *g*, the particular or *group* factors that he had identified, the *specific* factors or Spearman's *s* (attributes of a single trait measured on all occasions), and what Burt called *accidental* factors, or attributes of a single trait measured only on a single occasion.* Burt had synthesized all perspectives. In Spearman's terms, his theory was monarchic in recognizing the domination of *g*, oligarchic in its identification of group factors, and anarchic in recognizing *s*-factors for each test. But Burt's scheme was no compromise; it was Spearman's hierarchical theory with yet another level subordinate to *g*.

*This accidental variance, representing peculiarities of particular testing situations, forms part of what statisticians call "measurement error." It is important to quantify, for it may form a basic level of comparison for the identification of causes in a family of techniques called the "analysis of variance." But it represents the peculiarity of an occasion, not a quality either of a test or a testee.

Moreover, Burt accepted and greatly elaborated Spearman's views on the differential innateness of levels. Spearman had regarded *g* as inherited, *s* as a function of training. Burt agreed, but promoted the influence of education to his group factors as well. He retained the distinction between an inherited and ineluctable *g*, and a set of more specialized abilities amenable to improvement by education:

Although defect in general intelligence inevitably places a definite limit to educational progress, defect in special intellectual abilities rarely does so (1937, p. 537).

Burt also declared, with his usual intensity and persistence, that the primary importance of factor analysis lay in its capacity for identifying inherited, permanent qualities:

From the very outset of my educational work it has seemed essential, not merely to show that a general factor underlies the cognitive group of mental activities, but also that this general factor (or some important component of it) is innate or permanent (1940, p. 57).

The search for factors thus becomes, to a great extent, an attempt to discover inborn potentialities, such as will permanently aid or limit the individual's behavior later on (1940, p. 230).

Burt on the reification of factors

Burt's view on reification, as Hearnshaw has noted with frustration (1979, p. 166), are inconsistent and even contradictory (sometimes within the same publication).^{*} Often, Burt branded reification of factors as a temptation to be avoided:

No doubt, this causal language, which we all to some extent favor, arises partly from the irrepressible disposition of the human mind to reify and even to personify whatever it can—to picture inferred reasons as realities and to endow those realities with an active force (1940, p. 66).

^{*}Other scholars often complained of Burt's tendency to obfuscate, temporize, and argue both sides as his own when treating difficult and controversial issues. D. F. Vincent wrote of his correspondence with Burt about the history of factor analysis (in Hearnshaw, 1979, pp. 177–178): "I should not get a simple answer to a simple question. I should get half a dozen foolscap sheets of typescript, all very polite and very cordial, raising half a dozen subsidiary issues in which I was not particularly interested, and to which out of politeness I should have to reply . . . I should then get more foolscap pages of typescript raising more extraneous issues. . . . After the first letter my problem has been how to terminate the correspondence without being discourteous."

He spoke with eloquence about this error of thought:

The ordinary mind loves to reduce patterns to single atomlike existents—to treat memory as an elementary faculty lodged in a phrenological organ, to squeeze all consciousness into the pineal gland, to call a dozen different complaints rheumatic and regard them all as the effect of a specific germ, to declare that strength resides in the hair or in the blood, to treat beauty as an elementary quality that can be laid on like so much varnish. But the whole trend of current science is to seek its unifying principles, not in simple unitary causes, but in the system or structural pattern as such (1940, p. 237).

And he explicitly denied that factors were things in the head (1937, p. 459):

The "factors," in short, are to be regarded as convenient mathematical abstractions, not as concrete mental "faculties," lodged in separate "organs" of the brain.

What could be more clearly stated?

Yet in a biographical comment, Burt (1961, p. 53) centered his argument with Spearman not on the issue of whether or not factors should be reified, but rather *how* they should be reified: "Spearman himself identified the general factor with 'cerebral energy.' I identified it with the general structure of the brain." In the same article, he provided more details of suspected physical locations for entities identified by mathematical factors. Group factors, he argues, are definite areas of the cerebral cortex (1961, p. 57), while the general factor represents the amount and complexity of cortical tissue: "It is this general character of the individual's brain-tissue—viz., the general degree of systematic complexity in the neuronal architecture—that seems to me to represent the general factor, and account for the high positive correlations obtained between various cognitive tests" (1961, pp. 57–58; see also 1959, p. 106).^{*}

^{*}One might resolve this apparent contradiction by arguing that Burt refused to reify on the basis of mathematical evidence alone (in 1940), but did so later when independent neurological information confirmed the existence of structures in the brain that could be identified with factors. It is true that Burt advanced some neurological arguments (1961, p. 57, for example) in comparing the brains of normal individuals and "low grade defectives." But these arguments are sporadic, perfunctory, and peripheral. Burt repeated them virtually verbatim, in publication after publication, without citing sources or providing any specific reason for allying mathematical factors with cortical properties.

Lest one be tempted to regard these later statements as a shift in belief from the caution of a scholar in 1940 to the poor judgment of a man mired in the frauds of his later years, I note that Burt presented the same arguments for reification in 1940, right alongside the warnings against it:

Now, although I do not identify the general factor *g* with any form of energy, I should be ready to grant it quite as much "real existence" as physical energy can justifiably claim (1940, p. 214). Intelligence I regard not indeed as designating a special form of energy, but rather as specifying certain individual differences in the structure of the central nervous system—differences whose concrete nature could be described in histological terms (1940, pp. 216–217).

Burt even went so far as to suggest that the all-or-none character of neural discharge "supports the demand for an ultimate analysis into independent or 'orthogonal' factors" (1940, p. 222).

But perhaps the best indication of Burt's hope for reification lies in the very title he chose for his major book of 1940. He called it *The Factors of the Mind*.

Burt followed Spearman in trying to find a physical location in the brain for mathematical factors extracted from the correlation matrix of mental tests. But Burt also went further, and established himself as a reifier in a domain that Spearman himself would never have dared to enter. Burt could not be satisfied with something so vulgar and material as a bit of neural tissue for the residence of factors. He had a wider vision that evoked the spirit of Plato himself. Material objects on earth are immediate and imperfect representations of higher essences in an ideal world beyond our ken.

Burt subjected many kinds of data to factor analysis during his long career. His interpretations of factors display a Platonic belief in a higher reality, embodied imperfectly by material objects, but discernible in them through an idealization of their essential, underlying properties on principal component factors. He analyzed a suite of emotional traits (1940, pp. 406–408) and identified his first principal component as a factor of "general emotionality." (He also found two bipolar factors for extrovert-introvert and euphoric-sorrowful.) He discovered "a general paranormal factor" in a study of ESP data (in Hearnshaw, 1979, p. 222). He analyzed human anatomy and interpreted the first principal component as an ideal type for humanity (1940, p. 113).

One needn't, from these examples, infer Burt's belief in a literal, higher reality: perhaps he thought of these idealized general factors as mere principles of classification to aid human understanding. But, in a factor analysis of aesthetic judgment, Burt explicitly expressed his conviction that real standards of beauty exist, independent of the presence of human beings to appreciate them. Burt selected fifty postcards with illustrations ranging from the great masters down to "the crudest and most flashy birthday card that I could find at a paper shop in the slums." He asked a group of subjects to rank the cards in order of beauty and performed a factor analysis of correlations among the ranks. Again, he discerned an underlying general factor on the first principal component, declared it to be a universal standard of beauty, and expressed a personal contempt for Victorian ceremonial statuary in identifying this higher reality:

We see beauty because it is there to be seen. . . . I am tempted to contend that aesthetic relations, like logical relations, have an independent, objective existence: the Venus of Milo would remain more lovely than Queen Victoria's statue in the Mall, the Taj Mahal than the Albert Memorial, though every man and woman in the world were killed by a passing comet's gas.

In analyses of intelligence, Burt often claimed (1939, 1940, 1949, for example) that each level of his hierarchical, four-factor theory corresponded with a recognized category in "the traditional logic of classes" (1939, p. 85)—the general factor to the *genus*, group factors to *species*, specific factors to the *proprium*, and accidental factors to the *accidens*. He seemed to regard these categories as more than conveniences for human ordering of the world's complexity, but as necessary ways of parsing a hierarchically structured reality.

Burt certainly believed in realms of existence beyond the material reality of everyday objects. He accepted much of the data of parapsychology and postulated an oversoul or psychon—"a kind of group mind formed by the subconscious telepathic interaction of the minds of certain persons now living, together perhaps with the psychic reservoir out of which the minds of individuals now deceased were formed, and into which they were reabsorbed on the death of their bodies" (Burt quoted in Hearnshaw, 1979, p. 225). In this higher realm of psychic reality, the "factors of the

mind" may have real existence as modes of truly universal thought.

Burt managed to espouse three contradictory views about the nature of factors: mathematical abstractions for human convenience; real entities lodged in physical properties of the brain; and real categories of thought in a higher, hierarchically organized realm of psychic reality. Spearman had not been very daring as a reifier; he never ventured beyond the Aristotelian urge for locating idealized abstractions within physical bodies themselves. Burt, at least in part, soared beyond into a Platonic realm above and beyond physical bodies. In this sense, Burt was the boldest, and literally most extensive, reifier of them all.

Burt and the political uses of g

Factor analysis is usually performed on the correlation matrix of tests. Burt pioneered an "inverted" form of factor analysis, mathematically equivalent to the usual style, but based on correlation between persons rather than tests. If each vector in the usual style (technically called R-mode analysis) represents the scores of several people on a single test, then each vector in Burt's inverted style (called Q-mode analysis) reflects the results of several tests for a single person. In other words, each vector now represents a person rather than a test, and the correlation between vectors measures the degree of relationship between individuals.

Why did Burt go to such lengths to develop a technique mathematically equivalent to the usual form, and generally more cumbersome and expensive to apply (since an experimental design almost always includes more people than tests)? The answer lies in Burt's uncommon focus of interest. Spearman, and most other factorists, wished to learn about the nature of thought or the structure of mind by studying correlations between tests measuring different aspects of mental functioning. Cyril Burt, as official psychologist of the London County Council (1913-1932), was interested in ranking pupils. Burt wrote in an autobiographical statement (1961, p. 56): "[Sir Godfrey] Thomson was interested primarily in the description of the *abilities* tested and in the differences between those abilities; I was interested rather in the *persons* tested and in the differences between them" (Burt's italics).

Comparison, for Burt, was no abstract issue. He wished to assess pupils in his own characteristic way, based upon two guiding

principles: first (the theme of this chapter) that general intelligence is a single, measurable entity (Spearman's *g*); second (Burt's own *idée fixe*) that a person's general intelligence is almost entirely innate and unchangeable. Thus, Burt sought the relationship among persons in a *unilinear ranking of inherited mental worth*. He used factor analysis to validate this single scale and to plant people upon it. "The very object of the factor-analysis," he wrote (1940, p. 136), "is to deduce from an empirical set of test measurements a single figure for each single individual." Burt sought (1940, p. 176) "one ideal order, acting as a general factor, common to every examiner and to every examinee, predominating over, though no doubt disturbed by, other irrelevant influences."

Burt's vision of a single ranking based on inherited ability fueled the major political triumph in Britain of hereditarian theories of mental testing. If the Immigration Restriction Act of 1924 signalled the chief victory of American hereditarians in psychology, then the so-called examination at 11+ awarded their British counterparts a triumph of equal impact. Under this system for streaming children into different secondary schools, pupils took an extensive examination at age ten or eleven. As a result of these tests, largely an attempt to assess Spearman's *g* for each child, 20 percent were sent to "grammar" schools where they might prepare for entry to a university, while 80 percent were relegated to technical or "secondary modern" schools and regarded as unfit for higher education.

Cyril Burt defended this separation as a wise step for "warding off the ultimate decline and fall that has overtaken each of the great civilizations of the past" (1959, p. 117):

It is essential in the interests alike of the children themselves and of the nation as a whole, that those who possess the highest ability—the cleverest of the clever—should be identified as accurately as possible. Of the methods hitherto tried out the so-called 11+ exam has proved to be by far the most trustworthy (1959, p. 117).

Burt's only complaint (1959, p. 32) was that the test and subsequent selection came too late in a child's life.

The system of examination at 11+ and subsequent separation of schools arose in conjunction with a series of official reports issued by government committees during twenty years (the Hadow

reports of 1926 and 1931, the Spens report of 1938, the Norwood report of 1943, and the Board of Education's White Paper on Educational Reconstruction—all leading to the Butler Education Act of 1944, which set policy until the mid-1960s when the Labour party vowed to end selection at 11 plus). In the flak surrounding the initial revelation of Burt's fraudulent work, he was often identified as the architect of the 11+ examination. This is not accurate; Burt was not even a member of the various reporting committees, though he did consult frequently with them and he did write extensively for their reports.* Yet it hardly matters whether or not Burt's hand actually moved the pen. The reports embody a particular view of education, clearly identified with the British school of factor analysis, and evidently linked most closely with Cyril Burt's version.

The 11+ examination was an embodiment of Spearman's hierarchical theory of intelligence, with its innate general factor pervading all cognitive activity. One critic referred to the series of reports as "hymns of praise to the 'g' factor" (in Hearnshaw, 1979, p. 112). The first Hadow report defined intellectual capacity measured by tests in Burt's favored terms as i.g.c. (innate, general, cognitive) ability: "During childhood, intellectual development progresses as if it were governed largely by a single, central factor, usually known as 'general intelligence,' which may be broadly defined as *innate, all round, intellectual* [my italics for i.g.c.] ability, and appears to enter into everything the child attempts to think, say, or do: this seems the most important factor in determining his work in the classroom."

The 11+ owed its general rationale to the British factorists; in addition, several of its details can also be traced to Burt's school. Why, for example, testing and separation at age eleven? There were practical and historical reasons to be sure; eleven was about the traditional age for transition between primary and secondary schools. But the factorists supplied two important theoretical supports. First, studies on the growth of children showed that *g* varied

*Hearnshaw (1979) reports that Burt had greatest influence over the 1938 Spens report, which recommended sorting at 11 plus and explicitly rejected comprehensive schooling under a single roof thereafter. Burt was piqued at the Norwood report because it downgraded psychological evidence; but, as Hearnshaw notes, this annoyance "masked a basic agreement with the recommendations, which in principle did not differ so much from those of the Spens committee, which he had earlier approved."

widely in early life and first stabilized at about age eleven. Spearman wrote in 1927 (p. 367): "If once, then, a child of 11 years or so has had his relative amount of *g* measured in a really accurate manner, the hope of teachers and parents that he will ever rise to a much higher standing as a late-bloomer would seem to be illusory." Second, Burt's "group factors," which (for purposes of separation by general mental worth) could only be viewed as disturbers of *g*, did not strongly affect a child until after age eleven. The 1931 Hadow report proclaimed that "special abilities rarely reveal themselves in any notable degree before the age of 11."

Burt often claimed that his primary goal in supporting 11+ was a "liberal" one—to provide access to higher education for disadvantaged children whose innate talents might otherwise not be recognized. I do not doubt that a few children of high ability were thus aided, though Burt himself did not believe that many people of high intelligence lay hidden in the lower classes. (He also believed that their numbers were rapidly decreasing as intelligent people moved up the social ladder leaving the lower classes more and more depleted of intellectual talent—1946, p. 15. R. Herrnstein [1971] caused quite a ruckus with the identical argument, recycled, a few years back.)*

Yet the major effect of 11+, in terms of human lives and hopes, surely lay with its primary numerical result—80 percent branded as unfit for higher education by reason of low innate intellectual ability. Two incidents come to mind, memories of two years spent in Britain during the regime of 11+: children, already labeled sufficiently by the location of their school, daily walking through the streets of Leeds in their academic uniforms, readily identified by all as the ones who hadn't qualified; a friend who had failed 11+ but reached the university anyway because she had learned Latin on her own, when her secondary modern school did not teach it and universities still required it for entrance into certain courses (how many other working-class teenagers would have had the means or motivation, whatever their talents and desires?).

Burt was committed to his eugenic vision of saving Britain by finding and educating its few people of eminent talent. For the rest, I assume that he wished them well and hoped to match their education with their ability as he perceived it. But the 80 percent

*The recycling reached full and lengthy fruition when Herrnstein and Charles Murray used the same claim as the opening gambit and general basis for *The Bell Curve* (1994).

were not included in his plan for the preservation of British greatness. Of them, he wrote (1959, p. 123):

It should be an essential part of the child's education to teach him how to face a possible beating on the 11+ (or any other examination), just as he should learn to take a beating in a half-mile race, or in a bout with boxing gloves, or a football match with a rival school.

Could Burt feel the pain of hopes dashed by biological proclamation if he was willing seriously to compare a permanent brand of intellectual inferiority with the loss of a single footrace?

L. L. Thurstone and the vectors of mind

Thurstone's critique and reconstruction

L. L. Thurstone was born (1887) and bred in Chicago (Ph.D., University of Chicago, 1917, professor of psychology at his alma mater from 1924 to his death in 1955). Perhaps it is not surprising that a man who wrote his major work from the heart of America during the Great Depression should have been the exterminating angel of Spearman's g . One could easily construct a moral fable in the heroic mold: Thurstone, free from the blinding dogmas of class bias, sees through the error of reification and hereditarian assumptions to unmask g as logically fallacious, scientifically worthless, and morally ambiguous. But our complex world grants validity to few such tales, and this one is as false and empty as most in its genre. Thurstone did undo g for some of the reasons cited above, but not because he acknowledged the deeper conceptual errors that had engendered it. In fact, Thurstone disliked g because he felt that it was not real enough!

Thurstone did not doubt that factor analysis should seek, as its primary objective, to identify real aspects of mind that could be linked to definite causes. Cyril Burt named his major book *The Factors of the Mind*, Thurstone, who invented the geometrical depiction of tests and factors as vectors (Figs. 6.6, 6.7), called his major work (1935) *The Vectors of Mind*. "The object of factor analysis," Thurstone wrote (1935, p. 53), "is to discover the mental faculties."

Thurstone argued that Spearman and Burt's method of principal components had failed to identify true vectors of mind because it placed factor axes in the wrong geometrical positions.

He objected strenuously both to the first principal component (which produced Spearman's g) and to the subsequent components (which identified "group factors" in clusters of positive and negative projections of tests).

The first principal component, Spearman's g , is a grand average of all tests in matrices of positive correlation coefficients, where all vectors must point in the same general direction (Fig. 6.4). What psychological meaning can such an axis have, Thurstone asked, if its position depends upon the tests included, and shifts drastically from one battery of tests to another?

Consider Fig. 6.10 taken from Thurstone's expansion (1947) of the *Vectors of Mind*. The curved lines form a spherical triangle on the surface of a sphere. Each vector radiates from the center of the sphere (not shown) and intersects the sphere's surface at a point represented by one of the twelve small circles. Thurstone assumes that the twelve vectors represent tests for three "real" faculties of mind, A, B, and C (call them verbal, numerical, and spatial, if you will). The left set of twelve tests includes eight that primarily measure spatial ability and fall near C; two tests measure verbal ability and lie near A, while two reflect numerical skill. But there is nothing sacrosanct about either the number or distribution of tests in a battery. Such decisions are arbitrary; in fact, a tester usually can't impose a decision at all because he doesn't know, in advance, which tests measure what underlying faculty. Another battery of tests (right side of Fig. 6.10) may happen to include eight for verbal skills and only two each for numerical and spatial ability.

The three faculties, Thurstone believes, are real and invariant in position no matter how many tests measure them in any battery. But look what happens to Spearman's g . It is simply the average of all tests, and its position—the x in Fig. 6.10—shifts markedly for the arbitrary reason that one battery includes more spatial tests (forcing g near spatial pole C) and the other more verbal tests (moving g near verbal pole A). What possible psychological meaning can g have if it is only an average, buffeted about by changes in the number of tests for different abilities? Thurstone wrote of g (1940, p. 208):

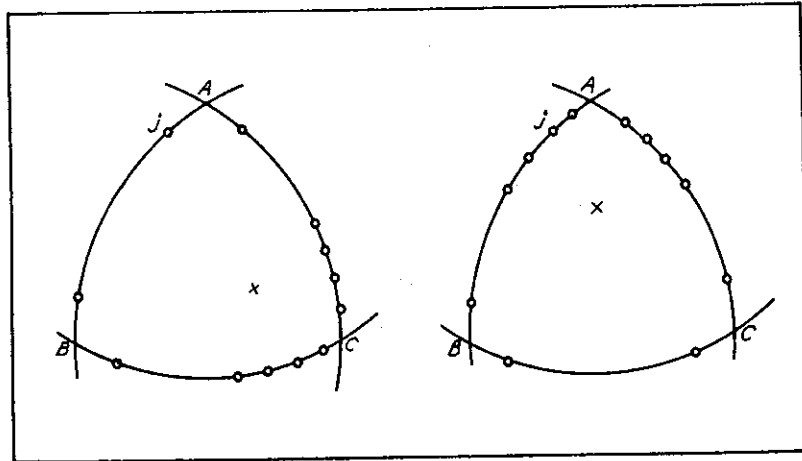
Such a factor can always be found routinely for any set of positively correlated tests, and it means nothing more or less than the average of all the abilities called for by the battery as a whole. Consequently, it varies

from one battery to another and has no fundamental psychological significance beyond the arbitrary collection of tests that anyone happens to put together. . . . We cannot be interested in a general factor which is only the average of any random collection of tests.

Burt had identified group factors by looking for clusters of positive and negative projections on the second and subsequent principal components. Thurstone objected strenuously to this method, not on mathematical grounds, but because he felt that tests could not have negative projections upon real "things." If a factor represented a true vector of *mind*, then an individual test might either measure that entity in part, and have a positive projection upon the factor, or it might not measure it at all, and have a zero projection. But a test could not have a negative projection upon a real vector of mind:

A negative entry . . . would have to be interpreted to mean that the possession of an ability has a detrimental effect on the test performance. One can readily understand how the possession of a certain ability can aid

6•10 Thurstone's illustration of how the position of the first principal component (the *x* in both figures) is affected by the types of tests included in a battery.



in a test performance, and one can imagine that an ability has no effect on a test performance, but it is difficult to think of abilities that are as often detrimental as helpful in the test performances. Surely, the correct factor matrix for cognitive tests does not have many negative entries, and preferably it should have none at all (1940, pp. 193-194).

Thurstone therefore set out to find the "correct factor matrix" by eliminating negative projections of tests upon axes and making all projections either positive or zero. The principal component axes of Spearman and Burt could not accomplish this because they, perforce, contained all positive projections on the first axis (*g*) and combinations of negative and positive groups on the subsequent "bipolars."

Thurstone's solution was ingenious and represents the most strikingly original, yet simple, idea in the history of factor analysis. Instead of making the first axis a grand average of all vectors and letting the others encompass a steadily decreasing amount of remaining information in the vectors, why not try to place all axes near clusters of vectors. The clusters may reflect real "vectors of mind," imperfectly measured by several tests. A factor axis placed near such a cluster will have high positive projections for tests measuring that primary ability* and very low zero projections for all tests measuring other primary abilities—as long as the primary abilities are independent and uncorrelated.

But how, mathematically, can factor axes be placed near clusters? Here, Thurstone had his great insight. The principal component axes of Burt and Spearman (Fig. 6.6) do not lie in the only position that factor axes can assume. They represent one possible solution, dictated by Spearman's a priori conviction that a single general intelligence exists. They are, in other words, theory-bound, not mathematically necessary—and the theory may be wrong. Thurstone decided to keep one feature of the Spearman-Burt scheme: his factor axes would remain mutually perpendicular, and therefore mathematically uncorrelated. The real vectors of mind, Thurstone reasoned, must represent *independent* primary abilities.

*Thurstone reified his factors, calling them "primary abilities," or "vectors of mind." All these terms represent the same mathematical object in Thurstone's system—factor axes placed near clusters of test vectors.

Thurstone therefore calculated the Spearman-Burt principal components and then *rotated* them to different positions until they lay as close as they could (while still remaining perpendicular) to actual clusters of vectors. In this rotated position, each factor axis would receive high positive projections for the few vectors clustered near it, and zero or near zero projections for all other vectors. When each vector has a high projection on one factor axis and zero or near zero projections on all others, Thurstone referred to the result as a *simple structure*. He redefined the factor problem as a search for simple structure by rotating factor axes from their principal components orientation to positions maximally close to clusters of vectors.

Figs. 6.6 and 6.7 show this process geometrically. The vectors are arranged in two clusters representing verbal and mathematical tests. In Fig. 6.6 the first principal component (g) is an average of all vectors, while the second is a bipolar, with verbal tests projecting negatively and arithmetic tests positively. But the verbal and arithmetic clusters are not well defined on this bipolar factor because most of their information has already been projected upon g , and little remains for distinction on the second axis. But if the axes are rotated to Thurstone's simple structure (Fig. 6.7), then both clusters are well defined because each is near a factor axis. The arithmetic tests project high on the first simple structure axis and low on the second; the verbal tests project high on the second and low on the first.

The factor problem is not solved pictorially, but by calculation. Thurstone used several mathematical criteria for discovering simple structure. One, still in common use, is called "varimax," or the search for *maximum variance* upon each rotated factor axis. The "variance" of an axis is measured by the spread of test projections upon it. Variance is low on the first principal component because all tests have about the same positive projection, and the spread is limited. But variance is high on rotated axes placed near clusters, because such axes have a few very high projections and other zero or near zero projections, thus maximizing the spread.*

The principal component and simple structure solutions are

*Readers who have done factor analysis for a course on statistics or methodology in the biological or social sciences will remember something about rotating axes to varimax positions. Like me, they are probably taught this procedure as if it were a mathematical deduction based on the inadequacy of principal components in find-

mathematically equivalent; neither is "better." Information is neither gained nor lost by rotating axes; it is merely redistributed. Preferences depend upon the meaning assigned to factor axes. The first principal component demonstrably exists. For Spearman, it is to be cherished as a measure of innate general intelligence. For Thurstone, it is a meaningless average of an arbitrary battery of tests, devoid of psychological significance, and calculated only as an intermediary step in rotation to simple structure.

Not all sets of vectors have a definable "simple structure." A random array without clusters cannot be fit by a set of factors, each with a few high projections and a larger number of near zero projections. The discovery of a simple structure implies that vectors are grouped into clusters, and that clusters are relatively independent of each other. Thurstone continually found simple structure among vectors of mental tests and therefore proclaimed that the tests measure a small number of independent "primary mental abilities," or vectors of mind—a return, in a sense, to an older "faculty psychology" that viewed the mind as a congeries of independent abilities.

Now it happens, over and over again, that when a factor matrix is found with a very large number of zero entries, the negative entries disappear at the same time. It does not seem as if all this could happen by chance. The reason is probably to be found in the underlying distinct mental processes that are involved in the different tasks. . . . These are what I have called primary mental abilities (1940, p. 194).

Thurstone believed that he had discovered real mental entities with fixed geometric positions. The primary mental abilities (or PMA's as he called them) do not shift their position or change their number in different batteries of tests. The verbal PMA exists in its designated spot whether it is measured by just three tests in one battery, or by twenty-five different tests in another.

The factorial methods have for their object to isolate the primary abilities by objective experimental procedures so that it may be a question of fact how many abilities are represented in a set of tasks (1938, p. 1).

ing clusters. In fact, it arose historically with reference to a definite theory of intelligence (Thurstone's belief in independent primary mental abilities) and in opposition to another (general intelligence and hierarchy of lesser factors) buttressed by principal components.

Thurstone reified his simple structure axes as primary mental abilities and sought to specify their number. His opinion shifted as he found new PMA's or condensed others, but his basic model included seven PMA's—V for verbal comprehension, W for word fluency, N for number (computational), S for spatial visualization, M for associative memory, P for perceptual speed, and R for reasoning.*

But what had happened to *g*—Spearman's ineluctable, innate, general intelligence—amidst all this rotation of axes? It had simply disappeared. It had been rotated away; it was not there anymore (Fig. 6.7). Thurstone studied the same data used by Spearman and Burt to discover *g*. But now, instead of a hierarchy with a dominant, innate, general intelligence and some subsidiary, trainable group factors, the same data had yielded a set of independent and equally important PMA's, with no hierarchy and no dominant general factor. What psychological meaning could *g* claim if it represented but one possible rendering of information subject to radically different, but mathematically equivalent, interpretations? Thurstone wrote of his most famous empirical study (1938, p. vii):

So far in our work we have not found the general factor of Spearman. . . . As far as we can determine at present, the tests that have been supposed to be saturated with the general common factor divide their variance among primary factors that are not present in all the tests. We cannot report any general common factor in the battery of 56 tests that have been analyzed in the present study.

The egalitarian interpretation of PMA's

Group factors for specialized abilities have had an interesting odyssey in the history of factor analysis. In Spearman's system they were called "disturbers" of the tetrad equation, and were often purposely eliminated by tossing out all but one test in a cluster—a remarkable way of rendering a hypothesis impervious to disproof. In a famous study, done specifically to discover whether or not

*Thurstone, like Burt, submitted many other sets of data to factor analysis. Burt, chained to his hierarchical model, always found a dominant general factor and subsidiary bipolars, whether he studied anatomical, parapsychological, or aesthetic data. Thurstone, wedded to his model, always discovered independent primary factors. In 1950, for example, he submitted tests of temperament to factor analysis and found primary factors, again seven in number. He named them activity, impulsiveness, emotional stability, sociability, athletic interest, ascendance, and reflectiveness.

group factors existed, Brown and Stephenson (1933) gave twenty-two cognitive tests to three hundred ten-year-old boys. They calculated some disturbingly high tetrads and dropped two tests "because 20 is a sufficiently large number for our present purpose." They then eliminated another for the large tetrads that it generated, excusing themselves by stating: "at worst it is no sin to omit one test from a battery of so many." More high values prompted the further excision of all tetrads including the correlation between two of the nineteen remaining tests, since "the mean of all tetrads involving this correlation is more than 5 times the probable error." Finally, with about one-fourth of the tetrads gone, the remaining eleven thousand formed a distribution close enough to normal. Spearman's "theory of two factors," they proclaimed, "satisfactorily passes the test of experience." "There is in the proof the foundation and development of a scientific experimental psychology; and, although we would be modest, to that extent it constitutes a 'Copernican revolution'" (Brown and Stephenson, 1933, p. 353).

For Cyril Burt, the group factors, although real and important in vocational guidance, were subsidiary to a dominant and innate *g*.

For Thurstone, the old group factors became primary mental abilities. They were the irreducible mental entities; *g* was a delusion.

Copernicus's heliocentric theory can be viewed as a purely mathematical hypothesis, offering a simpler representation for the same astronomical data that Ptolemy had explained by putting the earth at the center of things. Indeed, Copernicus's cautious and practical supporters, including the author of the preface to *De Revolutionibus*, urged just such a pragmatic course in a world populated with inquisitions and indices of forbidden books. But Copernicus's theory eventually produced a furor when its supporters, led by Galileo, insisted upon viewing it as a statement about the real organization of the heavens, not merely as a simpler numerical representation of planetary motion.

So it was with the Spearman-Burt vs. the Thurstone school of factor analysis. Their mathematical representations were equivalent and equally worthy of support. The debate reached a fury of intensity because the two mathematical schools advanced radically

different views about the real nature of intelligence—and the acceptance of one or the other entailed a set of fundamental consequences for the practice of education.

With Spearman's *g*, each child can be ranked on a single scale of innate intelligence; all else is subsidiary. General ability can be measured early in life and children can be sorted according to their intellectual promise (as in the 11+ examination).

With Thurstone's PMA's, there is no general ability to measure. Some children are good at some things, others excel in different and independent qualities of mind. Moreover, once the hegemony of *g* was broken, PMA's could bloom like the flowers in spring. Thurstone recognized only a few, but other influential schemes advocated 120 (Guilford, 1956) or perhaps more (Guilford, 1959, p. 477). (Guilford's 120 factors are not induced empirically, but predicted from a theoretical model—represented as a cube of dimensions $6 \times 5 \times 4 = 120$ —designating factors for empirical studies to find).

Unilinear ranking of pupils has no place, even in Thurstone's world of just a few PMA's. The essence of each child becomes his individuality, Thurstone wrote (1935, p. 53):

Even if each individual can be described in terms of a limited number of independent reference abilities, it is still possible for every person to be different from every other person in the world. Each person might be described in terms of his standard scores in a limited number of independent abilities. The number of permutations of these scores would probably be sufficient to guarantee the retention of individualities.

From the midst of an economic depression that reduced many of its intellectual elite to poverty, an America with egalitarian ideals (however rarely practiced) challenged Britain's traditional equation of social class with innate worth. Spearman's *g* had been rotated away, and general mental worth evaporated with it.

One could read the debate between Burt and Thurstone as a mathematical argument about the location of factor axes. This would be as myopic as interpreting the struggle between Galileo and the Church as an argument between two mathematically equivalent schemes for describing planetary motion. Burt certainly understood this larger context when he defended the 11+ examination against Thurstone's assault:

In educational practice the rash assumption that the general factor has at length been demolished has done much to sanction the impracticable idea that, in classifying children according to their varying capabilities, we need no longer consider their degree of general ability, and have only to allot them to schools of different types according to their special aptitudes; in short, that the examination at 11 plus can best be run on the principle of the caucus-race in Wonderland, where everybody wins and each get some kind of prize (1955, p. 165).

Thurstone, for his part, lobbied hard, producing arguments (and alternate tests) to support his belief that children should not be judged by a single number. He wished, instead, to assess each person as an individual with strengths and weaknesses according to his scores on an array of PMA's (as evidence of his success in altering the practice of testing in the United States, see Guilford, 1959, and Tuddenham, 1962, p. 515).

Instead of attempting to describe each individual's mental endowment by a single index such as a mental age or an intelligence quotient, it is preferable to describe him in terms of a profile of all the primary factors which are known to be significant. . . . If anyone insists on having a single index such as an I.Q., it can be obtained by taking an average of all the known abilities. But such an index tends so to blur the description of each man that his mental assets and limitations are buried in the single index (1946, p. 110).

Two pages later, Thurstone explicitly links his abstract theory of intelligence with preferred social views.

This work is consistent not only with the scientific object of identifying the distinguishable mental functions but it seems to be consistent also with the desire to differentiate our treatment of people by recognizing every person in terms of the mental and physical assets which make him unique as an individual (1946, p. 112).

Thurstone produced his fundamental reconstruction without attacking either of the deeper assumptions that had motivated Spearman and Burt—reification and hereditarianism. He worked within established traditions of argument in factor analysis, and reconstructed results and their meaning without altering the premises.

Thurstone never doubted that his PMA's were entities with identifiable causes (see his early work of 1924, pp. 146–147, for the

seeds of commitment to reifying abstract concepts—gregariousness in this case—as things within us). He even suspected that his mathematical methods would identify attributes of mind before biology attained the tools to verify them: "It is quite likely that the primary mental abilities will be fairly well isolated by the factorial methods before they are verified by the methods of neurology or genetics. Eventually the results of the several methods of investigating the same phenomena must agree" (1938, p. 2).

The vectors of mind are real, but their causes may be complex and multifarious. Thurstone admitted a strong potential influence for environment, but he emphasized inborn biology:

Some of the factors may turn out to be defined by endocrinological effects. Others may be defined by biochemical or biophysical parameters of the body fluids or of the central nervous system. Other factors may be defined by neurological or vascular relations in some anatomical locus; still others may involve parameters in the dynamics of the autonomic nervous system; still others may be defined in terms of experience and schooling (1947, p. 57).

Thurstone attacked the environmentalist school, citing evidence from studies of identical twins for the inheritance of PMA's. He also claimed that training would usually enhance innate differences, even while raising the accomplishments of both poorly and well-endowed children:

Inheritance plays an important part in determining mental performance. It is my own conviction that the arguments of the environmentalists are too much based on sentimentalism. They are often even fanatic on this subject. If the facts support the genetic interpretation, then the accusation of being undemocratic must not be hurled at the biologists. If anyone is undemocratic on this issue, it must be Mother Nature. To the question whether the mental abilities can be trained, the affirmative answer seems to be the only one that makes sense. On the other hand, if two boys who differ markedly in visualizing ability, for example, are given the same amount of training with this type of thinking, I am afraid that they will differ even more at the end of the training than they did at the start (1946, p. 111).

As I have emphasized throughout this book, no simple equation can be made between social preference and biological commitment. We can tell no cardboard tale of hereditarian baddies relegating whole races, classes, and sexes to permanent biological

inferiority—or of environmentalist goodies extolling the irreducible worth of all human beings. Other biases must be factored (pardon the vernacular usage) into a complex equation. Hereditarianism becomes an instrument for assigning groups to inferiority only when combined with a belief in ranking and differential worth. Burt united both views in his hereditarian synthesis. Thurstone exceeded Burt in his commitment to a naïve form of reification, and he did not oppose hereditarian claims (though he certainly never pursued them with the single-minded vigor of a Burt). But he chose not to rank and weigh on a single scale of general merit, and his destruction of Burt's primary instrument of ranking—Spearman's *g*—altered the history of mental testing.

Spearman and Burt react

When Thurstone dispersed *g* as an illusion, Spearman was still alive and pugnacious as ever, while Burt was at the height of his powers and influence. Spearman, who had deftly defended *g* for thirty years by incorporating critics within his flexible system, realized that Thurstone could not be so accommodated:

Hitherto all such attacks on it [*g*] appear to have eventually weakened into mere attempts to explain it more simply. Now, however, there has arisen a very different crisis; in a recent study, nothing has been found to explain; the general factor has just vanished. Moreover, the said study is no ordinary one. Alike for eminence of the author, for judiciousness of plan, and for comprehensiveness of scope, it would be hard to find any match for the very recent work on Primary Mental Abilities by L. L. Thurstone (Spearman, 1939, p. 78).

Spearman admitted that *g*, as an average among tests, could vary in position from battery to battery. But he held that its wandering was minor in scope, and that it always pointed in the same general direction, determined by the pervasive positive correlation between tests. Thurstone had not eliminated *g*; he had merely obscured it by a mathematical dodge, distributing it by bits and pieces among a set of group factors: "The new operation consisted essentially in scattering *g* among such numerous group factors, that the fragment assigned to each separately became too small to be noticeable" (1939, p. 14).

Spearman then turned Thurstone's favorite argument against him. As a convinced reifier, Thurstone believed that PMA's were

"out there" in fixed positions within a factorial space. He argued that Spearman and Burt's factors were not "real" because they varied in number and position among different batteries of tests. Spearman retorted that Thurstone's PMA's were also artifacts of chosen tests, not invariant vectors of mind. A PMA could be created simply by constructing a series of redundant tests that would measure the same thing several times, and establish a tight cluster of vectors. Similarly, any PMA could be dispersed by reducing or eliminating the tests that measure it. PMA's are not invariant locations present before anyone ever invented tests to identify them; they are products of the tests themselves:

We are led to the view that group factors, far from constituting a small number of sharply cut "primary" abilities, are endless in number, indefinitely varying in scope, and even unstable in existence. Any constituent of ability can become a group factor. Any can cease being so (1939, p. 15).

Spearman had reason to complain. Two years later, for example, Thurstone found a new PMA that he could not interpret (in Thurstone and Thurstone, 1941). He called it X_1 and identified it by strong correlations between three tests that involved the counting of dots. He even admitted that he would have missed X_1 entirely, had his battery included but one test of dotting:

All these tests have a factor in common; but since the three dot-counting tests are practically isolated from the rest of the battery and without any saturation on the number factor, we have very little to suggest the nature of the factor. It is, no doubt, the sort of function that would ordinarily be lost in the specific variance of the tests if only one of these dot-counting tests had been included in the battery (Thurstone and Thurstone, 1941, pp. 23-24).

Thurstone's attachment to reification blinded him to an obvious alternative. He assumed that X_1 really existed and that he had previously missed it by never including enough tests for its recognition. But suppose that X_1 is a creation of the tests, now "discovered" only because three redundant measures yield a cluster of vectors (and a potential PMA), whereas one different test can only be viewed as an oddball.

There is a general flaw in Thurstone's argument that PMA's are not test-dependent, and that the same factors will appear in any properly constituted battery. Thurstone claimed that an indi-

vidual test would always record the same PMA's only in simple structures that are "complete and overdetermined" (1947, p. 363)—in other words, only when all the vectors of mind have been properly identified and situated. Indeed, if there *really are* only a few vectors of mind, and if we can know when all have been identified, then any additional test must fall into its proper and unchanging position within the invariant simple structure. But there may be no such thing as an "overdetermined" simple structure, in which all possible factor axes have been discovered. Perhaps the factor axes are not fixed in number, but subject to unlimited increase as new tests are added. Perhaps they are truly test-dependent, and not real underlying entities at all. The very fact that estimates for the number of primary abilities have ranged from Thurstone's 7 or so to Guilford's 120 or more indicates that vectors of mind may be figments of mind.

If Spearman attacked Thurstone by supporting his beloved g , then Burt parried by defending a subject equally close to his heart—the identification of group factors by clusters of positive and negative projections on bipolar axes. Thurstone had attacked Spearman and Burt by agreeing that factors must be reified, but disparaging the English method for doing so. He dismissed Spearman's g as too variable in position, and rejected Burt's bipolar factors because "negative abilities" cannot exist. Burt replied, quite properly, that Thurstone was too unsubtle a reifier. Factors are not material objects in the head, but principles of classification that order reality. (Burt often argued the contrary position as well—see pp. 318-322.) Classification proceeds by logical dichotomy and antithesis (Burt, 1939). Negative projections do not imply that a person has less than zero of a definite thing. They only record a relative contrast between two abstract qualities of thought. More of something usually goes with less of another—administrative work and scholarly productivity, for example.

As their trump card, both Spearman and Burt argued that Thurstone had not produced a cogent revision of their reality, but only an alternative mathematics for the same data.

We may, of course, invent methods of factorial research that will always yield a factor-pattern showing some degree of "hierarchical" formation of (if we prefer) what is sometimes called "simple structure." But again the results will mean little or nothing: using the former, we could almost

always demonstrate that a general factor exists; using the latter, we could almost always demonstrate, even with the same set of data, that it does not exist (Burt, 1940, pp. 27-28).

But didn't Burt and Spearman understand that this very defense constituted their own undoing as well as Thurstone's? They were right, undeniably right. Thurstone had not proven an alternate reality. He had begun from different assumptions about the structure of mind and invented a mathematical scheme more in accord with his preferences. But the same criticism applies with equal force to Spearman and Burt. They too had started with an assumption about the nature of intelligence and had devised a mathematical system to buttress it. If the same data can be fit into two such different mathematical schemes, how can we say with assurance that one represents reality and the other a diversionary tinkering? Perhaps both views of reality are wrong, and their mutual failure lies in their common error: a shared belief in the reification of factors.

Copernicus was right, even though acceptable tables of planetary positions can be generated from Ptolemy's system. Burt and Spearman might be right even though Thurstone's mathematics treats the same data with equal facility. To vindicate either view, some legitimate appeal must be made outside the abstract mathematics itself. In this case, some biological grounding must be discovered. If biochemists had ever found Spearman's cerebral energy, if neurologists had ever mapped Thurstone's PMA's to definite areas of the cerebral cortex, then the basis for a preference might have been established. All combatants made appeals to biology and advanced tenuous claims, but no concrete tie has even been confirmed between any neurological object and a factor axis.

We are left only with the mathematics, and therefore cannot validate either system. Both are plagued with the conceptual error of reification. Factor analysis is a fine descriptive tool; I do not think that it will uncover the elusive (and illusory) factors, or vectors, of mind. Thurstone dethroned *g* not by being right with his alternate system, but by being equally wrong—and thus exposing the methodological errors of the entire enterprise.*

*Tuddenham (1962, p. 516) writes: "Test constructors will continue to employ factorial procedures, provided they pay off in improving the efficiency and predictive value of our test batteries, but the hope that factor analysis can supply a short inven-

Oblique axes and second-order g

Since Thurstone pioneered the geometrical representation of tests as vectors, it is surprising that he didn't immediately grasp a technical deficiency in his analysis. If tests are positively correlated, then all vectors must form a set in which no two are separated by an angle of more than 90° (for a right angle implies a correlation coefficient of zero). Thurstone wished to put his simple structure axes as near as possible to clusters within the total set of vectors. Yet he insisted that axes be perpendicular to each other. This criterion guarantees that axes cannot lie really close to clusters of vectors—as Fig. 6.11 indicates. For the maximal separation of vectors is less than 90° , and any two axes, forced to be perpendicular, must therefore lie outside the clusters themselves. Why not abandon this criterion, let the axes themselves be correlated (separated by an angle of less than 90°), and permit them to lie right within the clusters of vectors?

Perpendicular axes have a great conceptual advantage. They are mathematically independent (uncorrelated). If one wishes to identify factor axes as "primary mental abilities," perhaps they had best be uncorrelated—for if factor axes are themselves correlated, then doesn't the cause of that correlation become more "primary" than the factors themselves? But correlated axes also have a different kind of conceptual advantage: they can be placed nearer to clusters of vectors that may represent "mental abilities." You can't have it both ways for sets of vectors drawn from a matrix of positive correlation coefficients: factors may be independent and only close to clusters, or correlated and within clusters. (Neither system is "better"; each has its advantages in certain circumstances. Correlated and uncorrelated axes are both still used, and the argument continues, even in these days of computerized sophistication in factor analysis.)

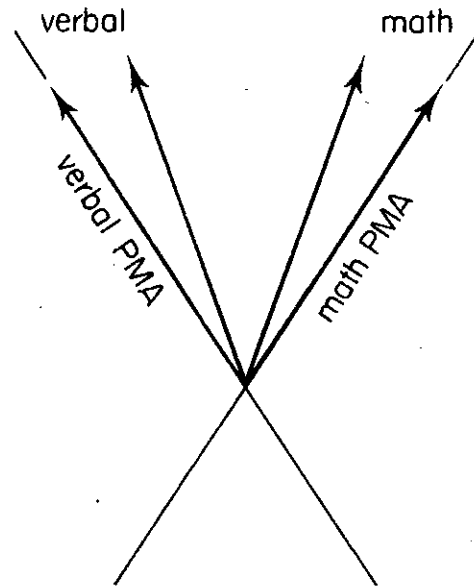
Thurstone invented rotated axes and simple structure in the early 1930s. In the late 1930s he began to experiment with so-

tory of 'basic abilities' is already waning. The continuous difficulties with factor analysis over the last half century suggest that there may be something fundamentally wrong with models which conceptualize intelligence in terms of a finite number of linear dimensions. To the statistician's dictum that whatever exists can be measured, the factorist has added the assumption that whatever can be 'measured' must exist. But the relation may not be reversible, and the assumption may be false."

called oblique simple structures, or systems of correlated axes. (Uncorrelated axes are called "orthogonal" or mutually perpendicular; correlated axes are "oblique" because the angle between them is less than 90° .) Just as several methods may be used for determining orthogonal simple structure, oblique axes can be calculated in a variety of ways, though the object is always to place axes within clusters of vectors. In one relatively simple method, shown in Fig. 6.11, actual vectors occupying extreme positions within the total set are used as factor axes. Note, in contrasting Figs. 6.7 and 6.11, how the factor axes for verbal and mathematical skills have moved from outside the actual clusters (in the orthogonal solution) to the clusters themselves (in the oblique solution).

Most factor-analysts work upon the assumption that correlations may have causes and that factor axes may help us to identify them. If the factor axes are themselves correlated, why not apply

6•11 Thurstone's oblique simple structure axes for the same four mental tests depicted in Figs. 6-6 and 6-7. Factor axes are no longer perpendicular to each other. In this example, the factor axes coincide with the peripheral vectors of the cluster.



the same argument and ask whether this correlation reflects some higher or more basic cause? The oblique axes of a simple structure for mental tests are usually positively correlated (as in Fig. 6.11). May not the cause of this correlation be identified with Spearman's g ? Is the old general factor ineluctable after all?

Thurstone wrestled with what he called this "second-order" g . I confess that I do not understand why he wrestled so hard, unless the many years of working with orthogonal solutions had set his mind and rendered the concept too unfamiliar to accept at first. If anyone understood the geometrical representation of vectors, it was Thurstone. This representation guarantees that oblique axes will be positively correlated, and that a second-order general factor must therefore exist. Second-order g is merely a fancier way of acknowledging what the raw correlation coefficients show—that nearly all correlation coefficients between mental tests are positive.

In any case, Thurstone finally bowed to inevitability and admitted the existence of a second-order general factor. He once even described it in almost Spearmanian terms (1946, p. 110):

There seems to exist a large number of special abilities that can be identified as primary abilities by the factorial methods, and underlying these special abilities there seems to exist some central energizing factor which promotes the activity of all these special abilities.

It might appear as if all the sound and fury of Thurstone's debate with the British factorists ended in a kind of stately compromise, more favorable to Burt and Spearman, and placing poor Thurstone in the unenviable position of struggling to save face. If the correlation of oblique axes yields a second-order g , then weren't Spearman and Burt right all along in their fundamental insistence upon a general factor? Thurstone may have shown that group factors were more important than any British factorist had ever admitted, but hadn't the primacy of g reasserted itself?

Arthur Jensen (1979) presents such an interpretation, but it badly misrepresents the history of this debate. Second-order g did not unite the disparate schools of Thurstone and the British factorists; it did not even produce a substantial compromise on either side. After all, the quotes I cited from Thurstone on the futility of ranking by IQ and the necessity of constructing profiles based on primary mental abilities for each individual were written after he

had admitted the second-order general factor. The two schools were not united and Spearman's *g* was not vindicated for three basic reasons:

1. For Spearman and Burt, *g* cannot merely exist; it must dominate. The *hierarchical* view—with a controlling innate *g* and subsidiary trainable group factors—was fundamental for the British school. How else could unilinear ranking be supported? How else could the 11+ examination be defended? For this examination supposedly measured a controlling mental force that defined a child's general potential and shaped his entire intellectual future.

Thurstone admitted a second-order *g*, but he regarded it as *secondary* in importance to what he continued to call "primary" mental abilities. Quite apart from any psychological speculation, the basic mathematics certainly supports Thurstone's view. Second-order *g* (the correlation of oblique simple structure axes) rarely accounts for more than a small percentage of the total information in a matrix of tests. On the other hand, Spearman's *g* (the first principal component) often encompasses more than half the information. The entire psychological apparatus, and all the practical schemes, of the British school depended upon the preeminence of *g*, not its mere presence. When Thurstone revised *The Vectors of Mind* in 1947, after admitting a second-order general factor, he continued to contrast himself with the British factorists by arguing that his scheme treated group factors as primary and the second-order general factor as residual, while they extolled *g* and considered group factors as secondary.

2. The central reason for claiming that Thurstone's alternate view disproves the necessary reality of Spearman's *g* retains its full force. Thurstone derived his contrasting interpretation from the same data simply by placing factor axes in different locations. One could no longer move directly from the mathematics of factor axes to a psychological meaning.

In the absence of corroborative evidence from biology for one scheme or the other, how can one decide? Ultimately, however much a scientist hates to admit it, the decision becomes a matter of taste, or of prior preference based on personal or cultural biases. Spearman and Burt, as privileged citizens of class-conscious Britain, defended *g* and its linear ranking. Thurstone preferred individual profiles and numerous primary abilities. In an

unintentionally amusing aside, Thurstone once mused over the technical differences between Burt and himself, and decided that Burt's propensity for algebraic rather than geometrical representation of factors arose from his deficiency in the spatial PMA:

The configurational interpretations are evidently distasteful to Burt, for he does not have a single diagram in his text. Perhaps this is indicative of individual differences in imagery types which lead to differences in methods and interpretation among scientists (1947, p. ix).

3. Burt and Spearman based their psychological interpretation of factors on a belief that *g* was dominant *and* real—an innate, general intelligence, marking a person's essential nature. Thurstone's analysis permitted them, at best, a weak second-order *g*. But suppose they had prevailed and established the inevitability of a dominant *g*? Their argument still would have failed for a reason so basic that it passed everybody by. The problem resided in a logical error committed by all the great factorists I have discussed—the desire to reify factors as entities. In a curious way, the entire history that I have traced didn't matter. If Burt and Thurstone had never lived, if an entire profession had been permanently satisfied with Spearman's two-factor theory and had been singing the praises of its dominant *g* for three-quarters of a century since he proposed it, the flaw would be as glaring still.

The fact of pervasive positive correlation between mental tests must be among the most unsurprising major discoveries in the history of science. For positive correlation is the prediction of almost every contradictory theory about its potential cause, including both extreme views: pure hereditarianism (which Spearman and Burt came close to promulgating) and pure environmentalism (which no major thinker has ever been foolish enough to propose). In the first, people do jointly well or poorly on all sorts of tests because they are born either smart or stupid. In the second, they do jointly well or poorly because they either ate, read, learned, and lived in an enriched or a deprived fashion as children. Since both theories predict pervasive positive correlation, the fact of correlation itself can confirm neither. Since *g* is merely one elaborate way of expressing the correlations, its putative existence also says nothing about causes.

Thurstone on the uses of factor analysis

Thurstone sometimes advanced grandiose claims for the explanatory scope of his work. But he also possessed a streak of modesty that one never detects in Burt or Spearman. In reflective moments, he recognized that the choice of factor analysis as a method records the primitive state of knowledge in a field. Factor analysis is a brutally empirical technique, used when a discipline has no firmly established principles, but only a mass of crude data, and a hope that patterns of correlation might provide suggestions for further and more fruitful lines of inquiry. Thurstone wrote (1935, p. xi):

No one would think of investigating the fundamental laws of classical mechanics by correlational methods or by factor methods, because the laws of classical mechanics are already well known. If nothing were known about the law of falling bodies, it would be sensible to analyze, factorially, a great many attributes of objects that are dropped or thrown from an elevated point. It would then be discovered that one factor is heavily loaded with the time of fall and with the distance fallen but that this factor has a zero loading in the weight of the object. The usefulness of the factor methods will be at the borderline of science.

Nothing had changed when he revised *The Vectors of Mind* (1947, p. 56):

The exploratory nature of factor analysis is often not understood. Factor analysis has its principal usefulness at the borderline of science. . . . Factor analysis is useful, especially in those domains where basic and fruitful concepts are essentially lacking and where crucial experiments have been difficult to conceive. The new methods have a humble role. They enable us to make only the crudest first map of a new domain.

Note the common phrase—useful “at the borderline of science.” According to Thurstone, the decision to use factor analysis as a primary method implies a deep ignorance of principles and causes. That the three greatest factorists in psychology never got beyond these methods—despite all their lip service to neurology, endocrinology, and other potential ways of discovering an innate biology—proves how right Thurstone was. The tragedy of this tale is that the British hereditarians promoted an innatist interpretation of dominant *g* nonetheless, and thereby blunted the hopes of millions.

Epilogue: Arthur Jensen and the resurrection of Spearman's *g*

When I researched this chapter in 1979, I knew that the ghost of Spearman's *g* still haunted modern theories of intelligence. But I thought that its image was veiled, and its influence largely unrecognized. I hoped that a historical analysis of conceptual errors in its formulation and use might expose the hidden fallacies in some contemporary views of intelligence and IQ. I never expected to find a modern defense of IQ from an explicitly Spearmanian perspective.

But then America's best-known hereditarian, Arthur Jensen (1979) revealed himself as an unreconstructed Spearmanian, and centered an eight-hundred-page defense of IQ on the reality of *g*. More recently, Richard Herrnstein and Charles Murray also base their equally long *Bell Curve* (1994) on the same fallacy. I shall analyze Jensen's error here and *The Bell Curve's* version in the first two essays at the end of the book. History often cycles its errors.

Jensen performs most of his factor analyses in Spearman and Burt's preferred principal components orientation (though he is also willing to accept *g* in the form of Thurstone's correlation between oblique simple structure axes). Throughout the book, he names and reifies factors by the usual invalid appeal to mathematical pattern alone. We have *g*'s for general intelligence as well as *g*'s for general athletic ability (with subsidiary group factors for hand and arm strength, hand-eye coordination, and body balance).

Jensen explicitly defines intelligence as “the *g* factor of an indefinitely large and varied battery of mental tests” (p. 249). “We identify intelligence with *g*,” he states. “To the extent that a test orders individuals on *g*, it can be said to be a test of intelligence” (p. 224). IQ is our most effective test of intelligence because it projects so strongly upon the first principal component (*g*) in factor analyses of mental tests. Jensen reports (p. 219) that Full Scale IQ of the Wechsler adult scale correlates about 0.9 with *g*, while the 1937 Stanford-Binet projects about 0.8 upon a *g* that remains “highly stable over successive age levels” (while the few small group factors are not always present and tend to be unstable in any case).

Jensen proclaims the “ubiquity” of *g*, extending its scope into realms that might even have embarrassed Spearman himself. Jensen would not only rank people; he believes that all God's creatures

can be ordered on a *g* scale from amoebae at the bottom (p. 175) to extraterrestrial intelligences at the top (p. 248). I have not encountered such an explicit chain of being since last I read Kant's speculations about higher beings on Jupiter that bridge the gap between man and God.

Jensen has combined two of the oldest cultural prejudices of Western thought: the ladder of progress as a model for organizing life, and the reification of some abstract quality as a criterion for ranking. Jensen chooses "intelligence" and actually claims that the performance of invertebrates, fishes, and turtles on simple behavioral tests represents, in diminished form, the same essence that humans possess in greater abundance—namely *g*, reified as a measurable object. Evolution then becomes a march up the ladder to realms of more and more *g*.

As a paleontologist, I am astounded. Evolution forms a copiously branching bush, not a unilinear progressive sequence. Jensen speaks of "different levels of the phyletic scale—that is, earthworms, crabs, fishes, turtles, pigeons, rats, and monkeys." Doesn't he realize that modern earthworms and crabs are descendants of lineages that have evolved separately from vertebrates for more than 500 million years? They are not our ancestors; they are not even "lower" or less complicated than humans in any meaningful sense. They represent good solutions for their own way of life; they must not be judged by the hubristic notion that one peculiar primate forms a standard for all of life. As for vertebrates, "the turtle" is not, as Jensen claims, "phylogenetically higher than the fish." Turtles evolved much earlier than most modern fishes, and they exist as hundreds of species, while modern bony fishes include almost twenty thousand distinct kinds. What then is "the fish" and "the turtle"? Does Jensen really think that pigeon-rat-monkey-human represents an evolutionary sequence among warm-blooded vertebrates?

Jensen's caricature of evolution exposes his preference for unilinear ranking by implied worth. With such a perspective, *g* becomes almost irresistible, and Jensen uses it as a universal criterion of rank:

The common features of experimental tests developed by comparative psychologists that most clearly distinguish, say, chickens from dogs, dogs from monkeys, and monkeys from chimpanzees suggests that they are

roughly scalable along a *g* dimension . . . *g* can be viewed as an interspecies concept with a broad biological base culminating in the primates (p. 251).

Not satisfied with awarding *g* a real status as guardian of earthly ranks, Jensen would extend it throughout the universe, arguing that all conceivable intelligence must be measured by it:

The ubiquity of the concept of intelligence is clearly seen in discussions of the most culturally different beings one could well imagine—extraterrestrial life in the universe. . . . Can one easily imagine "intelligent" beings for whom there is no *g*, or whose *g* is qualitatively rather than quantitatively different from *g* as we know it (p. 248).

Jensen discusses Thurstone's work, but dismisses it as a criticism because Thurstone eventually admitted a second-order *g*. But Jensen has not recognized that if *g* is only a numerically weak, second-order effect, then it cannot support a claim that intelligence is a unitary, dominant entity of mental functioning. I think that Jensen senses his difficulty, because on one chart (p. 220) he calculates both classical *g* as a first principal component and then rotates all the factors (including *g*) to obtain a set of simple structure axes. Thus, he records the same thing twice for each test—*g* as a first principal component and the same information dispersed among simple structure axes—giving some tests a total information of more than 100 percent. Since big *g*'s appear in the same chart with large loadings on simple-structure axes, one might be falsely led to infer that *g* remains large even in simple-structure solutions.

Jensen is contemptuous of Thurstone's orthogonal simple structure, dismissing it as "flatly wrong" (p. 675) and as "scientifically an egregious error" (p. 258). Since he acknowledges that simple structure is mathematically equivalent to principal components, why the uncompromising rejection? It is wrong, Jensen argues, "not mathematically, but psychologically and scientifically" (p. 675) because "it artificially hides or submerges the large general factor" (p. 258) by rotating it away. Jensen has fallen into a vicious circle. He assumes a priori that *g* exists and that simple structure is wrong because it disperses *g*. But Thurstone developed the concept of simple structure largely to claim that *g* is a mathematical artifact. Thurstone wished to disperse *g* and succeeded; it is no disproof of his position to reiterate that he did so.

Jensen also uses *g* more specifically to buttress his claim that the average difference in IQ between whites and blacks records an

innate deficiency of intelligence among blacks. He cites the quotation on p. 271 as "Spearman's interesting hypothesis" that blacks score most poorly with respect to whites on tests strongly correlated with *g*:

This hypothesis is important to the study of test bias, because, if true, it means that the white-black difference in test scores is not mainly attributable to idiosyncratic cultural peculiarities in this or that test, but to a general factor that all the ability tests measure in common. A mean difference between populations that is related to one or more small group factors would seem to be explained more easily in terms of cultural differences than if the mean group difference is most closely related to a broad general factor common to a wide variety of tests (p. 535).

Here we see a reincarnation of the oldest argument in the Spearmanian tradition—the contrast between an innate dominant *g* and trainable group factors. But *g*, as I have shown, is neither clearly a thing, nor necessarily innate if a thing. Even if data existed to confirm Spearman's "interesting hypothesis," the results could not support Jensen's notion of ineluctable, innate difference.

I am grateful to Jensen for one thing: he has demonstrated by example that a reified Spearman's *g* is still the only promising justification for hereditarian theories of mean differences in IQ among human groups. *The Bell Curve* of Herrnstein and Murray (1994) has reinforced this poverty, indeed bankruptcy, of justification for the theory of unitary, rankable, innate, and effectively immutable intelligence—for these authors also ground their entire edifice on the fallacy of Spearman's *g*. The conceptual errors of reification have plagued *g* from the start, and Thurstone's critique remains as valid today as it was in the 1930s. Spearman's *g* is not an ineluctable entity; it represents one mathematical solution among many equivalent alternatives. The chimerical nature of *g* is the rotten core of Jensen's work, *The Bell Curve*, and of the entire hereditarian school.

A final thought

The tendency has always been strong to believe that whatever received a name must be an entity or being, having an independent existence of its own. And if no real entity answering to the name could be found, men did not for that reason suppose that none existed, but imagined that it was something peculiarly abstruse and mysterious.

JOHN STUART MILL

SEVEN

A Positive Conclusion

WALT WHITMAN, that great man of little brain (see p.124), advised us to "make much of negatives," and this book has heeded his words, some might say with a vengeance. While most of us can appreciate a cleansing broom, such an object rarely elicits much affection; it certainly produces no integration. But I do not regard this book as a negative exercise in debunking, offering nothing in return once the errors of biological determinism are exposed as social prejudice. I believe that we have much to learn about ourselves from the undeniable fact that we are evolved animals. This understanding cannot permeate through entrenched habits of thought that lead us to reify and rank—habits that arise within social contexts and support them in return. My message, as I hope to convey it at least, is strongly positive for three major reasons.

Debunking as positive science

The popular impression that disproof represents a negative side of science arises from a common, but erroneous, view of history. The idea of unilinear progress not only lies behind the racial rankings that I have criticized as social prejudice throughout this book; it also suggests a false concept of how science develops. In this view, any science begins in the nothingness of ignorance and moves toward truth by gathering more and more information, constructing theories as facts accumulate. In such a world, debunking would be primarily negative, for it would only shuck some rotten apples from the barrel of accumulating knowledge. But the barrel of theory is always full; sciences work with elaborated contexts for explaining facts from the very outset. Creationist biology was dead