Part I

INTELLIGENCE AND ITS MEASUREMENT

CHAPTER 1

History of Theories and Measurement of Intelligence

N. J. Mackintosh

It would be difficult to start measuring "intelligence" without at least some implicit or intuitive theory of what intelligence is, and from the earliest Greek philosophers to the present day, many writers have enunciated their ideas about the nature of intelligence (see Sternberg, 1990). For Plato, it was the love of learning – and the love of truth; St. Augustine, on the other hand, believed that superior intelligence might lead people away from God. Thomas Hobbes in *Leviathan* went into more detail, arguing that superior intelligence involved a quick wit and the ability to see similarities between different things, and differences between similar things (ideas that have certainly found their way into some modern intelligence tests).

Measurement, however, implies something further: No one would be interested in *measuring* people's intelligence unless they believed that people differ in intelligence. Many early writers did of course believe this. Homer's Odysseus, in contrast to the other heroes of the *Iliad* and *Odyssey*, is often described as clever, resourceful, wily, and quick-witted. But not all theorists shared this belief. Adam Smith in *The Wealth of Nations* argued that the division of labor was responsible not only for that wealth but also for the apparent differences in the talents of a philosopher and a street porter. And when Francis Galton published *Hereditary Genius* in 1869, in which he sought to prove that people differed in their natural abilities, his cousin Charles Darwin wrote to him: "You have made a convert of an opponent . . . for I have always maintained that, excepting fools, men do not differ in intellect, only in zeal and hard work" (Galton, 1908, p. 290).

Measuring Intelligence

Galton

Francis Galton had no doubt on this score.

> *I have no patience with the hypothesis occasionally expressed, and often implied, especially in tales written to teach children to be good, that babies are born pretty much alike, and that the sole agencies in creating differences between boy and boy, and man and man, are steady application and*

> *moral effort. It is in the most unqualified manner that I object to pretensions of natural equality. The experiences of the nursery, the school, the University, and of professional careers, are a chain of proofs to the contrary. (Galton, 1869, p. 12)*

The results of public examinations, he claimed, confirmed his belief. Even among undergraduates of Cambridge University, for example, there was an enormous range in the number of marks awarded in the honor examinations in mathematics, from less than 250 to over 7,500 in one particular two-year period. As a first (not entirely convincing) step in the development of his argument that this wide range of marks arose from variations in natural ability, he established that these scores (like other physical measurements) were normally distributed, the majority of candidates obtaining scores close to the average, with a regular and predictable decline in the proportion obtaining scores further away from the average.

Allied to an almost compulsive desire to measure anything and everything, it was perhaps inevitable that Galton should wish to provide a direct measure of such differences in natural ability. But what measures would succeed in doing this? In 1884, at the International Health Exhibition held in London, he set up an Anthropometric Laboratory, where for a small fee visitors could be measured for their keenness of sight and hearing, color vision, reaction time, manual strength, breathing power, height, weight and so on. He could hardly have supposed that these were all interchangeable measures of intelligence, and some were surely there simply because they could be measured. But Galton was a follower of the British empiricist philosophers and argued that if all knowledge comes through the senses, then a "larger," more intelligent mind must be one capable of finer sensory discrimination and thus able to store and act upon more sensory information. Hence the relation between intelligence and discrimination – which we will come across again.

J. McK. Cattell

A more systematic attempt to measure differences in mental abilities was proposed by James McKeen Cattell (1890), who published a detailed list of 10 "mental tests" (plus another 40 in brief outline); they included measures of two-point tactile threshold, just noticeable difference for weights, judgment of temporal intervals, reaction time, and letter span. Cattell did not claim that this rather heterogeneous collection of tests would provide a good measure of intelligence – indeed the word "intelligence" does not even appear in his paper. Once again, it seems clear that the tests were chosen largely because the techniques required were already available. These were the standard experimental paradigms of the new experimental psychology being developed in Germany, and whatever it was that they were measuring, at least one could hope that they were measuring it accurately. Although no doubt unfair, it is hard to resist the analogy with the man who has lost his keys when out at night, and confines his search to an area underneath a street lamp, not because he thinks that is where he lost them, but because at least he can see there.

As a measure of intelligence, indeed, Cattell's tests did not last long. Their demise came from a study conducted in his laboratory by Wissler (1901), who administered the tests to undergraduates at Columbia University and reported two seemingly devastating findings. First, although the students did indeed differ in their performance on many of the tests, there was virtually no correlation between their performance on one and their performance on another. Even the correlations between different measures of speed, for example, averaged less than .20. If one test, therefore, was succeeding in measuring differences in intelligence, the others could not be. But which was the successful one? The second finding suggested that none of them were, for there was essentially no correlation between any of the tests and the students' college grades, which did in fact tend to correlate with one another, and which, following Galton, presumably were

reflecting differences in intellectual ability between the students.

Binet

It was the Frenchman, Alfred Binet, who solved the problem of devising an apparently satisfactory measure of intelligence. Although he and his colleague, Victor Henri, had made earlier attempts to measure differences in intelligence, they had not been spectacularly successful (Binet & Henri, 1896), and it was a commission from the French Ministry of Education that revived their efforts. The introduction of (nearly) universal primary education had brought into elementary schools a number of children of apparently below average intelligence, who would never had attended school before. They did not seem to be profiting from normal classroom teaching and were deemed to be in need of special education. The problem was to devise a quick and inexpensive way of identifying such children. Binet had little time for the new experimental psychology coming from Wundt's laboratory in Leipzig, and although much less hostile to the associationist tradition of British empiricism, he did not believe that associationism could answer all questions. Above all, he thought it nonsense to suppose that intelligence could be reduced to simple sensory function or reaction time. Observation of his own young daughters had convinced him that they were just as good as adults at making fine sensory discriminations, and although their average reaction time might be longer than that of an adult, this was not because they could never respond rapidly but rather because they occasionally responded very slowly – a failure Binet attributed (perhaps rather presciently as I shall show later) to lapses of attention.

For Binet, "intelligence" consisted in a multiplicity of different abilities and depended on a variety of "higher" psychological faculties – attention, memory, imagination, common sense, judgment, abstraction. Even more important, it involved coping successfully with the world and would thus be best measured by tests that required young children to show they were capable of coping with everyday problems. Could they follow simple instructions such as pointing to their nose and mouth? Did they understand the difference between morning and afternoon, and know what a fork is used for? Could they count the number of items in a display, and name the months of the year (in correct order)? And so on. Were these adequate measures of intelligence? Binet's critical insight was that as young children become more intellectually competent as they grow older, a good measure of intelligence would be one that older children found easier than younger ones; this was particularly relevant for his main task of identifying children who were mildly or perhaps more seriously retarded: The difference between "normal" and retarded children was that the former passed his tests at a younger age than the latter.

The validity of a particular item as a measure of intelligence in 6-year-old children, then, was that most children of this age could pass it, while essentially all 8-year-olds, but many fewer 4-year-olds, could. Thus Binet and his later collaborator Theodore Simon devised a series of different tests of increasing difficulty, for 4-, 6-, 8-, and 10-year-old children, all based on this empirical insight and extensive trial and error (Binet & Simon, 1908). They acknowledged that there was no abrupt cutoff to most children's performance. A normal 6-year-old would probably answer nearly all the items in the 4-year test, most of those in the 6-year test, but quite possibly also manage one or two in the 8-year test. It was only with some reluctance and in a later paper (Binet & Simon, 1911) that he was prepared to assign any precise score (a mental age) to an individual child.Stern (1912) later introduced the concept of the intelligence quotient or IQ, defined as mental age divided by chronological age, but he seems to have set little store by the innovation that has guaranteed his place in so many textbooks. He does not so much as mention it in his autobiography (Stern 1930). Binet's reluctance to provide any precise measurement of a child's

intelligence arose partly from his important observation that different children might get exactly the same total number of items in each test correct, but with quite different patterns of correct and incorrect answers. This simply confirmed his belief that "intelligence" involved a number of more or less independent faculties.

Spearman and the Theory of General Intelligence

Faculty psychology was Charles Spearman's bête noire. He abhorred the program that would separate the mind into a loose confederation of independent faculties of learning, memory, attention, and so on. What was needed was to understand its operations as a whole. Without knowing about Wissler's experiment, he repeated something very like it with a group of young children in a village school (Spearman, 1904; he later admitted that had he been aware of Wissler's results he would probably never have run his own study). He obtained independent ratings of each child's "cleverness in school" (from their teacher) and "sharpness and common sense out of school" (from two older children), and also measured their performance on three sensory tasks. Unlike Wissler, he did observe modest positive correlations between all his measures: the average correlation between the three ratings of intelligence was .55; that between the three sensory measures was .25, and that between the intelligence and sensory measures was .38. These were certainly more encouraging than Wissler's results – perhaps because the obvious restriction of range in students at Columbia University lowered Wissler's correlations. But they were still rather modest. Undaunted, Spearman argued that this was because his measures were *unreliable*, and a correction for attenuation had to be applied. The true correlation between two tests was the observed correlation between them divided by the square root of the product of their reliabilities. This is of course a standard formula for "disattenuating" correlations between two tests, but in modern test theory, the reliability of a test is measured by the correlation between performance on the test on separate occasions, or performance on one half of the test versus the other. Spearman had no such information and instead assumed that the reliability of his three measures of intelligence was the observed correlation between them, and similarly for the three sensory measures. Armed with this assumption, he was able to calculate the "true" correlation between intelligence and sensory discrimination:

$$r(\text{true}) = .38/\sqrt{(.55 \times .25)} = 1.01.$$

Of course, correlations cannot actually be greater than 1.0, but Spearman assumed that this was a minor error and confidently asserted that he had shown that general intelligence was general sensory discrimination.

In fact, Spearman later acknowledged that these measures of reliability were inappropriate, and he did not pursue the argument about the identity of intelligence and sensory discrimination. A much more important observation was one he made in data collected in another school, where he obtained somewhat more objective measures of academic performance, namely, each child's rank order in class for each of four different subjects, as well as measures of pitch discrimination and musical ability as rated by their music teacher. Interestingly, he anticipated Binet's appreciation of the importance of age by making an allowance for a pupil's age in adjusting their class ranking. The correlation matrix he reported between all these six measures is shown in Table 1.1. As can be seen, the correlations form what Spearman called a "hierarchy"; with one small exception, the correlations decrease as one goes down each column or across each row of the matrix. What was the meaning of this? Spearman's "Two Factor" theory provided the proposed answer. Each test measures its own specific factor, but also, to a greater or lesser extent, a general factor that is common to *all* the tests in the battery. It is this general factor, which Spearman labeled *g* for general intelligence,

Table 1.1. Spearman's reported correlations between six different measures of school attainment and musical performance. The figures comes from Spearman (1904) – although Fancher (1985), going back to Spearman's raw data, has shown that they are not, alas, perfectly accurate

	Classics	*French*	*English*	*Maths*	*Pitch*	*Music*
Classics	–					
French	.83	–				
English	.78	.67	–			
Maths	.70	.67	.64	–		
Pitch	.66	.65	.54	.45	–	
Music	.63	.57	.51	.51	.40	

that was said to explain why all tests correlated with one another. That this was a sufficient explanation of the observed correlation matrix, Spearman argued, was proved by the application of his "tetrad equation." If $r_{1.2}$ stands for the observed correlation between tests 1 and 2 and so on, then the tetrad equation was as follows:

$$r_{1.2} \times r_{3.4} = r_{1.3} \times r_{2.4} \quad (1)$$

Substitute the appropriate numbers from Table 1.1 into this equation, and you have $.83 \times .64 = .53$, and $.78 \times .67 = .52$, as close as one could reasonably ask – and much the same will hold for any other two pairs of correlations in the table. Why should this be? Spearman's explanation was straightforward: The reason that tests 1 and 2 correlate is because both measure *g*. The observed correlation between the two tests is simply a product of each test's separate correlation with *g*:

$$r_{1.2} = r_{1.g} \times r_{2.g} \quad (2)$$

And because this is true of all other pairs of tests, equation 1 can be rewritten as follows:

$$r_{1.g} \times r_{2.g} \times r_{3.g} \times r_{4.g} = r_{1.g} \times r_{3.g} \times r_{2.g} \times r_{4.g} \quad (3)$$

which is clearly true. When the correlation matrix of a battery of tests forms a hierarchy such as that seen in Table 1.1, to which the tetrad equation applies, the explanation, said Spearman, is because the correlations between all tests are entirely due to each test's correlation with the single general factor, *g*.

It is worth remarking that the development of Spearman's two-factor theory was not based on the results of anything that could properly be called an intelligence test. But that theory allowed Spearman later to argue that Binet's tests, without Binet's knowing it, had in fact succeeded in providing a good measure of general intelligence. Every item in Binet's tests measured its own specific factor as well as the general factor. Over the test as a whole, however, the specific factors would, so to say, cancel each other out, leaving the general factor to shine strongly through. This was the principle of "the indifference of the indicator." More or less any mental test battery, witheringly referred to as any "hotchpotch of multitudinous measurements" (Spearman, 1930, p. 324), would end up measuring general intelligence, provided only that it was sufficiently large and sufficiently diverse.

What was the explanation of the general factor? At different times, Spearman came up with two quite different explanations. One was couched in terms of his "noegenetic" laws, which asserted that the three fundaments of general intelligence were the apprehension of one's own experience, the eduction of relations and the eduction of correlates (Spearman, 1930). The second was that *g* was "something of the nature of an "energy" or "power" that serves in common

the whole cortex" (Spearman, 1923, p. 5). Two of the noegenetic laws bore fruit in that their emphasis on the importance of the perception of relations between superficially dissimilar items, otherwise known as analogical reasoning, provided the impetus for the construction of Raven's Matrices (Penrose & Raven, 1936). The second perhaps bears some passing resemblance to more modern ideas, discussed below, that speed of information processing is the basis of *g* (Anderson, 1992; Jensen, 1998).

The Divorce between Theory and Practice

Binet's tests were introduced into the United States by Henry Goddard, the director of research at the Vineland Training School in New Jersey, an institution for individuals with developmental disabilities. These tests later formed the initial basis for Lewis Terman's greatly improved version, the Stanford-Binet test (Terman, 1916), now in its fifth edition (Roid, 2003). Terman and Goddard then joined the committee set up by Robert Yerkes to devise the U.S. Army Alpha and Beta tests used to screen some 1.75 million draftees in World War I. The apparent success of these tests and the wide publicity they attracted after the war led to a proliferation of new test construction – with many new tests based on the Army tests themselves but most designed for use in schools, where they were often used to assign children to different tracks or classes. The first on the scene was the National Intelligence Test developed by Yerkes and Brigham, but later tests included the Henmon-Nelson tests, and the Otis "Quick Scoring Mental Ability Tests." For such tests to be economically viable, it was important that they could be administered to relatively large numbers of people in a relatively short time. In other words, they needed to be group tests, and as the name of the Otis test implies, one desideratum was that they could be rapidly and reliably scored. Hence the introduction of the multiple-choice question format. Brigham also developed tests for the College Entrance Exam Board, which were the forerunners of the Scholastic Aptitude Test (SAT). Eventually more individual tests were devised, including the first individual test of adult intelligence, the Wechsler-Bellevue test, the forerunner of the Wechsler Adult Intelligence Scale (WAIS), but which also borrowed and adapted many items from the Army tests. Wechsler also introduced the concept of the "deviation IQ." IQ defined as mental age divided by chronological age might work for children up to the age of 16 or so, but because 40-year-old adults do not obtain mental age scores twice those of 20-year-olds, mental ages will not work for adults. Wechsler's solution was to compare an individual's test score with the average score obtained by people of the same age.

Both Goddard and Terman had stressed the practical *usefulness* of Binet's test and Terman's revision of it. Goddard argued that the tests identified not only those referred to at that time as "idiots" and "imbeciles" – those severely disabled with an IQ score below 50 – but also, and even more important because they were not so easy to diagnose by other methods, the mildly disabled or "feebleminded" (for whom Goddard coined the term "moron"). Goddard (1914) had no doubt that it was in society's best interests to curb the reproduction of such individuals – and in this echoing eugenic views that were commonplace at the time (see Kevles, 1985) – but this association has served to give IQ tests a bad name ever since (e.g., Murdoch, 2007). Terman (1916), in his introduction to the Stanford-Binet test, also spent much time extolling the test's practical value, not only for identifying the "feebleminded" but also in schools, where much time would be saved by identifying the more and the less able. Later test constructors also stressed the value of identifying intellectually gifted children. The important point for the test constructors was to establish the predictive validity of their tests. Test scores would not only identify the disabled but also predict who would do well at school, who would therefore profitably continue on

to college and university, and thereafter who would be suitable for what job. Many organizations, including, for example, the military and the police, routinely gave all applicants an IQ test and imposed a lower cutoff score as a minimum admission requirement.

In sharp contrast to Binet, who regarded his tests as simply providing an estimate of a child's present level of intellectual functioning, Spearman, Burt, Goddard, Terman, and Yerkes were also united in their conviction that their tests "were originally intended, and are now definitely known, to measure native intellectual ability" (Yoakum & Yerkes, 1920, p. 27). It hardly needs to be said that they had not a shred of real evidence for this conviction. But it too did little to endear other psychologists to the psychometric tradition – especially when this hereditarian bias was combined with one that saw differences in average native ability between different social or racial groups.

All this contributed to the independent development of IQ tests as a technology, divorced from mainstream psychology, and, it is commonly assumed, without any theoretical understanding of the nature of the intelligence they were supposed to be measuring. But Galton and Binet both had theories of intelligence, and both supposed that a successful measure of intelligence would be guided by their theory. Wissler's results suggested that Galton's theory was wrong, while the success of Binet's test perhaps implies that his theory was right. The trouble was that although it was indeed based on some empirical observation of his children, it was a rather commonsensical theory that owed little to the experimental psychology of his day. Galton's and especially Cattell's ideas were indeed based on contemporary experimental psychology – but that psychology, in the shape of Wissler's data, had apparently shown they were wrong. This concatenation of events is often blamed for the development of the two separate disciplines of psychology, the experimental and the correlational, so famously lamented by Cronbach (1957).

This must be at least a large part of the story – but perhaps not quite all. In his autobiography, Spearman (1930, p. 326) had referred to the division between what he called general and individual psychology as "among the worst evils in modern psychology." He was not talking about Wissler's data in this context. The truth of the matter is surely that for much of the 20th century, and certainly in the early years of the century, experimental psychology had no worthwhile theory of intelligence or cognition to offer. Intelligence tests could not be based on a psychological theory of intelligence because there was no such theory. Neither Binet's nor Spearman's "theories" could really be said to provide a satisfactory explanation of what it is to be more or less intelligent. Any rapprochement between experimental and correlational psychology had to wait on the development of theory in cognitive psychology – and that did not happen until the final quarter of the century.

Factor Analysis

In the meantime, what was left for psychometricians to do? The answer was that they developed new intelligence tests and explored the relationships between them. One impetus for this was, as implied above, to cash in on the popularity of any measure that seemed to promise the practical advantages held out by Terman, Yerkes, and Brigham. A theoretically much more important rationale was to assess the adequacy of Spearman's two-factor theory: Would all test batteries yield a "hierarchy" consistent with the idea that all correlations between tests could be explained by postulating a single general factor? This was of course a theoretical question, and to that extent test developers were exploring theories of intelligence. The question was soon answered in the negative: A correlation matrix that reveals clusters of high correlations between some tests separated by lower correlations between these tests and another cluster of high correlations will disconfirm the tetrad equation. Burt (1917) claimed to find evidence of a cluster of high correlations between different "verbal" tests

while El Koussy (1935) found a similar cluster of high correlations between a variety of "spatial" tests. New techniques of factor analysis made clear the need to postulate additional "group factors" in addition to *g*. Then Thurstone (1938) argued that a different procedure for factor analysis (rotation to simple structure) eliminated the need for any *g* at all: Instead, there were a number of independent "primary mental abilities," suspiciously akin to Spearman's detested faculties. Thurstone identified seven in all, including verbal comprehension, verbal fluency, number, spatial visualization, inductive reasoning, memory, and possibly perceptual speed, and designed a series of tests, his Primary Mental Abilities (PMA) tests, that were intended to provide measures of each distinct ability.

In a separate development, Raymond Cattell proposed that Spearman's *g* should be divided into two distinct but correlated factors, fluid and crystallized intelligence, Gf and Gc, the former reflecting the ability to solve problems such as Raven's Matrices, the latter measured by tests of knowledge, such as vocabulary (Cattell, 1971; Horn & Cattell, 1966). In Cattell's original account, Gf was seen as the biological basis of intelligence, and Gc as the expression of that ability in the accumulated knowledge acquired as a result of exposure to a particular culture. That particular formulation of the theory was abandoned by Horn, who argued (surely correctly) that the ability to solve the analogical reasoning and series completion tasks that measure Gf are just as dependent on past learning (even if not explicitly taught in school) as are the tests of vocabulary or general knowledge that define Gc (see Horn & Hofer, 1992). Nevertheless, most modern accounts of the structure of intelligence have acknowledged the importance of the distinction between Gf and Gc. More to the point, at least one modern test battery, the W-J III (Woodcock-Johnson test) has been designed in part to provide separate measures of Gc and Gf – as well as of other components of intelligence identified by the theory.

It soon became apparent, and was acknowledged by Thurstone himself, that his primary mental abilities were not in fact wholly independent. The pervasive "positive manifold" reflected the fact that performance on any one test was correlated with performance on all other tests, and *g* reappeared to account for the correlation between Thurstone's primary abilities. As early as 1938, Holzinger and Harman (1938) had proposed one way of doing this, but the preferred method was later introduced by Schmid and Leiman (1957) in their "orthogonalized hierarchical" solution. In his magisterial survey of 20th century factorial studies, Carroll (1993) concluded that the structure of intellectual abilities revealed by factor analysis included a general factor, *g*, at a third "stratum," some half dozen or more broad group factors, including Gf and Gc at a second stratum, as well as factors of visuospatial abilities (Gv), retrieval (Gr), and processing speed (Gs), and a large, perhaps indefinite number of specific factors at a first stratum. This is now sometimes referred to as the Carroll-Horn-Cattell (or CHC) model and could be seen as a reconciliation between, or amalgamation of, Spearman's and Thurstone's accounts, the first and third strata corresponding to Spearman's general and specific factors, the second stratum to Thurstone's primary mental abilities.

The story does not, of course, end here. Other factorists, most famously Guilford (1967, 1985, 1988), in his structure-of-intellect model, postulated a far larger number of abilities than Thurstone had ever dreamed of. He started with 120, moved to 150 and ended up with 180; the novel feature of his account was that these abilities were derived from theoretical first principles: particular abilities were said to consist of five different kinds of operation, applied to five different types of content, expressed in terms of one of six different products (this produced the 150 number). Although initially skeptical of the need to postulate a higher order general factor, later versions of the model did include a general factor. Guilford's abilities should be seen as corresponding to the numerous specific first stratum abilities in the CHC model. One of the virtues of his approach is that he included

measures of creativity and social intelligence that have not commonly appeared in traditional IQ test batteries. Suss and Beauducel (2005) have provided a sympathetic account, and Brody a rather less sympathetic one which concluded that "Guilford's theory is without empirical support" (Brody, 1992, p. 34). There also remain those, such as Gould (1997) and Gardner (1993), who have disputed whether there is any general factor at all. Without going as far as Guilford, Gardner believes that there are eight or possibly more distinct intelligences, most of them not measured by IQ tests at all. He is surely right to suppose that traditional IQ tests fail to measure important aspects of human intelligence. But it seems merely perverse to deny, or seek to explain away, the fact that a general factor will be revealed by analysis of most batteries of mental tests. The pervasive positive manifold guarantees that a significant general factor will emerge from factor analysis of virtually any battery of cognitive tests – and this applies as strongly to tests of most of Gardner's intelligences as it does to traditional IQ test batteries (Visser, Ashton, & Vernon, 2006).

Within the more traditional mainstream, Johnson and Bouchard (2005) have rejected the factorial structure proposed by Carroll and Horn and Cattell in favor of one advanced by Vernon (1950), in which *g* sits above two group factors, *v:ed* and *k:m*, the former verbal-educational, the latter spatial and mechanical. They claimed that Vernon's structure, slightly modified, provided a better fit to two large datasets they analyzed than either Carroll's account or Horn and Cattell's Gf-Gc theory. In the Vernon model, fluid reasoning is part of *g* rather than identified as Gf, while *k:m* refers to perceptual and spatial abilities rather than more general reasoning. Vernon's *v:ed* is a specifically verbal ability, as opposed to Gc, which can include figural knowledge. It is surely too soon to pass judgment on this dispute.

Factor analysis has clearly had important implications for theories of human intelligence. Spearman and Thurstone initially held diametrically opposed views about the structure of abilities, and factor analysis of different test batteries eventually forced them both to acknowledge that their original theories had been wrong – even if each had also been partly right. So it would be quite wrong to claim that mainstream research on human intelligence was, for most of the 20th century, conducted in a theoretical void. But the theories in question were theories about the structure of human abilities and the relationship between different aspects or components of intelligence, not about the nature of the operations, processes, or mechanisms underlying these abilities. Factor analysis was never going to answer these questions.

What is *g*?

Although most intelligence researchers today probably accept that the general factor is here to stay, they remain sharply divided on its explanation. These disagreements go well beyond a rejection of Spearman's specific suggestions that *g* is either mental energy or the eduction of relations and correlates.

One of the earliest scholars to raise a much wider issue and to question the *logic* of Spearman's account of *g* was Thomson (1916), who argued that the positive manifold arises, not because all tests measure a single psychological or neurobiological process, as Spearman supposed, but because each test taps a subset of a very large number of elementary processes or operations, and there will almost necessarily be some overlap between the processes engaged by one test and those engaged by another. In general, if tests 1 and 2 each engage a proportion, P_1 and P_2, of the mind's elementary operations, the correlation between the two tests will be $\surd P_1 \times P_2$. There is no doubt that Thomson's argument is valid – although it has not been taken up in the form he presented it. But Ceci (1990) pointed out that the fact that three tests, 1, 2, and 3, all correlate with one another does not necessarily imply that there is any process common to all three. If each test depended on two processes, test 1 on a and b, test 2 on b and c, and