JAMES R. FLYNN

THE ONTOLOGY OF INTELLIGENCE

The purpose of this paper is to clarify the existential or ontological status of intelligence and this means an analysis of both the primitive concept of intelligence found in everyday life and scientific constructs. Scientific constructs can be understood only in the context of the theories that have generated them and I have chosen to concentrate on the Spearman—Jensen theory of intelligence and the psychometric construct called *g* or general intelligence. The first half of the paper will argue that psychometric *g* has shown considerable promise as a scientific construct; the second half will show that its successes have been accompanied by significant failures and argue that the Spearman—Jensen theory must be revised and transcended.

PSYCHOMETRIC *g* AND ITS CRITICS

The selection of the Spearman—Jensen theory reflects the personal view that A. R. Jensen has done most to give the concept of intelligence explanatory power. The presentation of the theory will cover the following topics: the relationship between the primitive concept of intelligence and IQ tests; the derivation of *g* from certain performance trends on IQ tests; the critical debate about *g*; the ontology and explanatory power of *g*. Throughout I will attempt to defend Jensen against certain of his critics, on the grounds that the theory has often been unfairly attacked and its merits unacknowledged.

Intelligence and IQ Tests

Measuring intelligence is a prerequisite to giving it explanatory power which immediately poses the question of whether IQ tests measure intelligence. Early on Jensen (1972, pp. 75—77) asserted that intelligence was by definition what IQ tests measure. Block and Dworkin (1977, p. 415) responded with a critique of this kind of crude operationalism. Any attempt to define something in terms of the readings given by a measuring instrument leads to absurdities. Defining tem-

John Forge (Ed.), Measurement, Realism and Objectivity, 1—40.

perature in terms of what thermometers measure denies the possibility of defective thermometers or the invention of a new instrument better than any now existent. Defining intelligence as what IQ tests measure would leave us unable to say any IQ test was better than another. Actually Jensen's so-called definition consists of a single sentence and was not really meant to define anything. It was a badly phrased attempt to introduce the construct *g* which is derived from performance trends on IQ tests.

The need for a concept of intelligence independent of IQ tests suggest an analysis of the primitive concept found in everyday life. Jensen (1979, pp. 80—81) argues that it arises in a comparative context. A Robinson Crusoe would become aware that he remembered things and learned things even in total isolation but would realize he was quicker than others at learning things only with a companion. Thus, we have the concept of intelligence as a mental ability distinct from memory and learning and indeed speak of people as intelligent even though they may be forgetful or ignorant. Perhaps this is why the concept has such a long history; that is, from ancient India to ancient Greece, from St. Thomas to Binet and Spearman, the notion persists that some people have 'better minds' than others and that this has to do with abstract problem-solving, induction and deduction, transfer of learning from one situation to another, the perception of relationships.

The later definitions of intelligence are not much improvement on the earlier ones and therefore the obvious next step is to move from the pre-scientific to a scientific concept, a measurable variable with explanatory power. However, this does not mean that the two concepts can proceed in complete isolation from one another. The primitive concept of intelligence describes a certain role that scientific concepts attempt to play: they are supposed to measure a mental ability with great importance for the life-histories of individuals and groups; whether someone is intelligent should tell us something about their academic and occupational achievements and groups that have a high average intelligence ought to show a high standard of cultural attainment. Moreover, if psychology discovered a potent mental ability distinct from memory and learning, and that ability happened to bear little resemblance to the primitive concept, it would be important to say so. People in general should not be allowed to think that those who score well or poorly on IQ tests are 'bright' or 'dull' if what the tests measure has little to do with the ordinary meaning of those words.

The case that IQ tests measure something close to the primitive concept of intelligence attempts to match test scores with popular assessments of intelligence and popular expectations about intelligent behaviour. For example, correlations between teacher rankings of pupils for intelligence and their tested IQs average at about 0.70; the meaning of a value like 0.70 will be explained shortly but for now, it should be taken as moderately significant. More impressive, the reasons for discrepancies between teachers and tests appear to favor the tests. Teachers tend to rank girls over boys, presumably influenced by their greater docility and application, they tend to rank extroverts over introverts, and they usually ignore age. The tests allow for age differences, that is, a 10-year-old who scores as well as an 11-year-old is given a higher IQ. The notion that intelligence increases with age during childhood is certainly not counter-intuitive; parents often cite precocious behaviour as evidence that a young child is bright (Jensen, 1980a, pp. 173—174).

The strongest piece of evidence comes from Terman's famous study of high IQ subjects (Terman and Oden, 1959). In 1921, he selected a sample of 1528 children as having IQs above 140 and their life histories are a good match for popular expectations about the real world achievements of the highly intelligent. As children they read much, had wide interests, and seven out of eight were ahead of their age group at school while none had been held back. As adults an extraordinary number had earned degrees, entered professions, achieved high positions, published books and articles, and had biographical citations in leading reference works. Children with IQs above 150 sometimes show quite remarkable ability and by the age of 12 or 13 are capable of excelling in university courses. Jensen points out that low IQ also tends to tally with what popular expectations would dictate. He reports an interview with a subject whose greatest interest was baseball (Jensen, 1981, p. 65). Despite regular attendance and watching telecasts, the low-IQ subject did not know the number of players, the names of the positions, or most of the rules.

High IQ is generally a prerequisite for high achievement in mathematics, science, and other areas dependent on academic learning, but it is not a sufficient condition. Aside from the obvious need for non-intellectual traits such as motivation, persistence, good health, there is clearly a special talent for things like mathematics and music, a creative imagination that IQ tests cannot identify. And outside academic areas, people can excel in commercial dealing, social skills, acting, and design without particularly high IQs. Correlation coefficients add confirmation: they are moderately high between IQ and standardized scholastic achievement tests, or between IQ and school grades, but quite low with things like income and job performance in most occupations (Block and Dworkin, 1977, pp. 439—444). This has caused Bane and Jencks (1977, p. 326) to conclude that IQ tests measure only one rather limited variety of intelligence. Whether one considers this kind of intelligence, call it academic intelligence, of limited importance depends on what one thinks of academic pursuits. I think it important to have people who can write and read serious works of philosophy, history, literature, mathematics, and science, even if the kind of intelligence they possess has little market value and says little about their social skills.

The contention that IQ tests measure something close to the primitive concept of intelligence has surrounded that concept with a number of limitations. However, most of these have analogues in popular opinion; that is, 'IQ' could be replaced by the word 'intelligence' in the last paragraph without doing much violence to ordinary language. We all know 'bright' people, particularly people with academic ability, who are poorly motivated, hopeless with numbers, socially obtuse, practically inept, and also people accomplished in the performing arts and organizational roles who do not seem particularly intelligent. Economos (1980, p. 342) gives perhaps the best summary of the case in favour of IQ tests: it proposes "that people who score poorly on these tests will almost always find it harder, for example, to follow advanced mathematical reasoning, or quickly to extract the meaning from a scholarly paragraph, than will people who score well on them"; but it does not say that these are the only or the most desirable or the most advantageous of human abilities.

The Construct Called g

Assume we could rank representative samples of a population on a number of non-team sports, golf, bowling, archery, shooting, the ten events of the decathalon with a long-distance event, say the half-marathon, thrown in to make a total of 15. It becomes apparent that the same people tend to be good or bad at most of the sports in question. That is, the variation in the percentile scores of one person across the 15 sports tends to be less than the score variance of 15 randomly selected people on any one sport. Therefore, we speak of general athletic performance, meaning for the moment nothing more than to describe the tendency towards consistent quality of performance across sports. This is of course something that impresses us in real life: people who beat us at practically everything, no matter how hard we try, people of whom we say, he or she is just a better athlete than I am.

Assume we invented a mathematical technique for measuring the tendency towards consistent performance across these 15 events or sports and, moreover, invented an artificial event, no more complicated than real events, on which a person's level of performance was an index of their general level of performance. Indeed, it predicts general athletic performance better than what might seem to be the obvious method: taking a sample from various sports, that is, a test that included one hole of golf, one sprint, one weight event, and so forth. Having operationalized the tendency towards consistent performance, we would be justified in calling it an ability. After all, a person's performance on this artificial event would tell us something about their potential performance across sports in general, including those they had never tried. We would therefore call it a measure of general athletic ability. This would not mean that it measured something unitary in the sense of being one simple skill. It would measure a functionally interrelated set of skills no more and no less complex than those used in a particular sport, say golf or the pole vault. At present, nothing would be known about the physiological prerequisites of good performance, or the causes, or how to enhance performance.

The trend towards consistency of performance is also manifest across a variety of mental tests and this despite great diversity of content. No matter whether tests feature vocabulary, general information, verbal oddities, scrambled sentences, logical reasoning, inferential conclusions, number series, pictorial oddities, spatial analogies, figure generalization, or completion of matrices, the same people tend to be good or bad at them. The construct called *g* is essentially a mathematical device that measures this tendency towards consistent performance across tests. The great psychologist Charles Spearman (1904) invented factor analysis and derived the first *g*, although initially from things that today would hardly qualify as mental tests. Rather than presenting a technical account of factor analysis, I will try to convey the logic behind how *g* is calculated. The following assumes that all tests have standard scores and show a bivariate or multivariate normal distribution; and that the population mean and standard deviation have been set at 100 and 15 respectively.

Another way of describing the above tendency is to say that all IQ tests have a positive correlation with one another and therefore, the best starting point is to explain what a correlation coefficient is. A correlation coefficient measures the slope of the regression line, for example, take two tests with a positive correlation of 0.70. If we used the first test to select out an elite group with a mean IQ of 110, that group would also be an elite on the second test with a mean IQ of 107. In other words, the correlation coefficient told us that a group 10 points above average on the first test would keep 70 per cent of their advantage on the second, that is, 7 points. If the tests had a perfect correlation of 1.00, they would keep all 10 points and remain steady at an IQ of 110. If the correlation was nil or 0.00, they would regress all the way to the population average and have an IQ of 100. Clearly the correlation coefficient is a measure of the tendency of those who do well on one test to excel on another.

Now *g* is essentially nothing more than a super correlation coefficient that measures the slope of the regression line when you have many tests, say 15 rather than just two. The *g* loading for any one test of the 15 tells how much an elite on that test will regress in terms of their overall performance on all 15 tests taken collectively. If the first test had a *g* loading of 0.70, an elite group with a mean IQ of 110 would score 107 on their overall performance; if the second test had 0.60, a similar elite would score 106; and so forth. The value for *g* itself represents the correlation between a test whose loading is typical of the 15 tests and overall performance on all 15 tests collectively. Jensen (undated, p. 19) has tried to approximate the value one would get for *g* from a total of more than 70 mental tests, deriving an estimate of 0.65. In other words, given a collection of all mental tests in common use, the typical regression would be from a performance of 110 on a particular test to a performance of 106.5 on all tests taken together. Clearly *g* measures a significant tendency of those who do well on one test to excel on all mental tests.

Indeed, the scores of one person across 15 tests tend to vary less than the scores of 15 randomly selected people on any one test. Jensen's estimate of the percentage of variance accounted for by *g* is 42.7 percent: just as taking the square root of variance accounted for will give the correlation coefficient, so the square root of 0.427 was used to derive the value of 0.65 for *g* noted above. We can also calculate the variance *not* accounted for by *g* as follows: 1.000 − 0.427 = 0.573; and 0.573 × 225 (total variance) = 129. So variance of performance across tests is only 129 compared to variance within a particular test of 225. Since the standard deviation is the square root of variance, the across test *SD* is 11.35 compared to a within test *SD* of 15. What this means in practice is this: take a group with a mean IQ of 115 on a particular mental test, one with a typical *g* loading. This performance would put them at the 84th percentile of the whole population on that test. If they then took a battery of 15 mental tests, they would tend to score below average or below the 50th percentile on only one or two tests. An elite on a particular test shows a strong tendency to remain an elite on other tests. Therefore we speak of general mental test performance, meaning for the moment nothing more than to describe the tendency towards consistent quality of performance across all mental tests.

During the 1930s, L. Penrose and J. C. Raven invented a new test, Raven's Progressive Matrices or Ravens for short, in an effort to maximize *g* loading. They largely succeeded, that is, a group that scores above or below the population mean on Ravens will hardly regress towards the mean at all in terms of their overall performance on a diverse collection of tests. Indeed, it predicts general performance, say on a collection of 15 tests, better than what might seem to be the obvious method: making up a composite test by random selection of items from each of the 15 tests! The fact that Ravens operationalizes the tendency towards consistent performance across mental tests raises this question: can we now move from saying *g* measures general mental test *performance* to saying it measures general mental test *ability*? I would answer in the affirmative, not merely because of what Ravens predicts but because of what it operationalizes.

The mere fact something predicts general mental test performance would not in itself encourage us to say that it measures an ability. Socioeconomic status also predicts test performance and it is a measure of things like income. It is the content of Ravens that is so impressive. Each item presents a pattern in which there is a gap, followed by six alternatives each of which would fit that gap like fitting a piece into a jig-saw puzzle. The subject must choose the piece with the correct markings, those markings which alone would render the total pattern complete. The patterns can be made very complex, with few or many elements, some shaded and some unshaded, some derived from others by rotation, flipping over, or as mirror images. In other words, Ravens consists entirely of perceptual analogy items based on geometrical patterns, items that require making comparisons, seeing similarities and differences, reasoning by analogy, and perceiving a consistent pattern amid irrelevant complexity. This fact plus the fact Ravens is such an excellent predictor of general mental test performance, the two in combination, make a *prima facie* case for the following: those who possess a relatively small number of functionally interrelated mental skills have a significant advantage on a huge diversity of mental test items. Moreover, these items have little apparent functional relationship with one another; once again they range from vocabulary to number series, general information to logical reasoning, verbal comprehension to figure generalization, verbal oddities to coding digits.

The fact that people who possess a set of functionally interrelated skills have an advantage over other people on a wide range of apparently unrelated tasks is all I mean when using the word 'ability'. Similarly we speak of someone with good timing, fast reflexes, and good coordination as possessing athletic ability across a wide range of sports that apparently are unrelated in terms of tasks. Therefore I am prepared to speak of general mental test ability. Thus far, that ability refers only to good performance on mental tests and possible links to the real world, both causal and consequential, remain to be discussed.

Cattell (1963) has made a start towards explaining how the ability in question could cause good performance on the tasks in question. He noted that mental tests with heavy *g* loadings divide themselves into two very different sorts: tests that have little informational content but demand the ability to see relationships between relatively simple elements, such as Ravens, which he called tests of *fluid g*; and tests that emphasize already acquired knowledge, such as vocabulary, general information, and arithmetic, which he called tests of *crystalized g*. The hypothesis is that Ravens and these other tests usually correlate so highly because a person with the ability Ravens measures will, given normal cultural opportunities, be the sort of person who acquires a large vocabulary, wide general information, and so forth. Indeed, when subjects from different backgrounds are tested, it is sometimes found that they can score equally well on Ravens despite differential vocabularies and stores of information, that is, they have the same level of ability but have applied it to different cultural raw material.

The growth curves of fluid and crystalized *g* are consistent with this hypothesis. Ravens performance improves throughout childhood, holds at a stable maximum between ages 18 and 25, then begins a gradual decline which accelerates after 60. Performance on vocabulary and general information tests can increase throughout life right up to 60 or 70 years of age and decline thereafter only gradually if at all. This certainly does not contradict observations from everyday life: that basic mental ability or agility declines after youth but that accumulated knowledge and its attendent skills can increase until old age.

Controversy About g

There has been much critical debate about *g* and while some of this is best postponed, certain objections are so fundamental that they must be addressed at once.

First, since *g* is essentially a correlation coefficient calculated from performance on a collection of mental tests, it will differ from one collection to another. Even were it possible to calculate it for all mental tests in common use, this collection alters over time as tests are added or discarded. Thurstone (1940, p. 208) asked how *g* could have any psychological significance given that it measures performance on "an arbitrary collection of tests anyone happens to put together", and that it alters radically if one first includes only spatial tests, then only verbal ones, then only numerical tests, and so forth. Jensen (1980a, pp. 233–234) answers that based on very high correlations, the *g* of one set of tests is very much the same as the *g* of another, just so long as the sets are both large and diverse. Each should include 10 or more tests and sample a wide range of information, tasks, and materials inclusive of verbal, figural, and numerical items. He stresses that the best contemporary IQ tests, such as the Wechsler, have 10 or more subtests and that these easily possess the required diversity.

The presupposition that lends strength to this answer is that such IQ tests measure a *g* which is a good bet to win the prize dangled before our eyes by the primitive concept of intelligence: that they measure a mental ability which may well have explanatory power when applied to the life-histories of individuals or groups. This can only be established by testing *g* repeatedly in a programme of scientific research. There is no problem about ensuring that *g* will remain the same throughout:

Ravens can be used as the marker test of fluid *g* and the Wechsler tests as markers when a mix of both fluid and crystalized *g* is required.

However, the above objection makes a point Jensen would not dispute: the quality of *g* will be a function of the quality of the tests from which it is derived. If we had only tests of school learning, general performance on these would give deceptive expectations about able people only recently exposed to good education. If we had only memory tests, general performance would not differentiate black and white Americans in the same way current IQ tests do. And if, for some reason, the best of present-day IQ tests were found wanting, then we would need both new tests and a new psychometric *g*.

A second objection: if *g* is test relative, are not the tests themselves culturally relative? This is undoubtedly to some degree the case: a modern industrial society dependent on science and technology emphasizes abstract problem-solving; a pre-industrial society may emphasize the memory needed to absorb oral tradition or the information-processing skills needed to survive in the bush. But it would be much to the credit of psychology if it could measure even those mental abilities needed in industrial society. Moreover, cultural relativism offers little to those groups some psychologists believe to be genetically inferior in terms of *g*. American blacks want to succeed in American society and if they are genetically handicapped for that, that is what counts. It is no solace to be told that they would have a genetic advantage were they still living in the environment of their African ancestors. As Lewontin (1977, p. 81) has pointed out, the argument of specific cultural origins of IQ testing cuts both ways.

A third objection is based on factor analysis, that is, the mathematical technique by which *g* is calculated. Gould (1981, pp. 310 and 314) emphasizes that factor analysis of performance on a collection of tests can be used to extract either the general factor called *g* or a multiplicity of other factors. He argues that since mathematics gives no guide, whether one calculates *g* or other factors becomes a matter of personal preference or bias. However, he grants that the scientific status of various factors could be tested by real-world data, such as evidence of a link to biological entities. Unless biology has a monopoly on the vindication of *g*, Gould's admission that scientific considerations can give a rational guide to our preferences about factors robs his objection of its force.

Factor analysis and performance trends on IQ tests do leave us with choices but these are not mutually exclusive. There is a strong tendency towards consistency of performance across IQ tests in general, but there is of course an even stronger consistency within subgroups of similar tests, that is, within the subgroup of purely verbal tests, or numerical tests, or spatial tests. Therefore, if a collection of 15 tests divides into three distinct subgroups, one can either calculate a general factor as a measure of the positive correlation of all the tests with one another, or a number of specific factors as a measure of the correlations within each kind of test, in this case a verbal factor, a numerical factor, and a spatial factor. In order to extract maximum scientific value from performance trends on IQ tests, it seems sensible to do both. The verbal factor may be the best predictor of marks in English courses and *g* the best predictor of results if students decide to broaden their education. As for seeking biological or physiological correlates, both general and specific factors might prove valuable.

Returning to sports, focusing on those who do well on a specific event, like distance running, has turned up something interesting. There is a high correlation between good performance and pulse rate, particularly how long it takes the pulse to return to normal after it has reached its maximum rate during vigorous exercise. This suggests investigating the efficiency of the cardio-vascular system as a factor in endurance. On the other hand, focusing on those who do well across a variety of sports and calculating a general factor could also prove valuable. The sporting dominance of American blacks has been so striking as to prompt investigation of physiological correlates: the major one discovered thus far is a faster conduction time from nerve to muscle as measured by electrodes. That is, there may be a correlation between general good performance in sport and the speed of stimulus-response or the reflex arc. This suggests investigating certain areas of body chemistry, such as the electrolyte or salt balance, and the roles of sweating and diet. The data is tentative and the status of general factors is not high at present in exercise physiology. But the point is that no-one can know in advance what factors will prove of scientific interest. Those who wish to investigate psychometric *g* have every right to invest their time in so doing.

The critics of a general factor often list the things it cannot do: ranking people by *g* does not tell us who will understand what ails a troubled friend; ranking students by *g* does little to diagnose their specific academic problems; ranking applicants by *g* is not a proper method of selecting a police force. As McClelland (1977, p. 58) points out, criterion sampling should be used to test applicants for most jobs. It creates tests based on analysis of what police actually do and should do, the tasks they perform, the vocabulary they use to communicate with the public, the fact they should not be racially biased. But it was a mistake to ever suppose a measure of general mental ability could do all these things. The primitive concept of intelligence hardly suggests that intelligence rankings would provide a detailed diagnosis of academic problems or an adequate criterion for selecting a police force. Psychometric *g* need not do everything in order to have its own unique scientific value.

A fourth objection contends that the positive correlations between IQ tests and between test items may appear synthetic but they are operationally analytic. As Block and Dworkin (1977, pp. 463—464) point out, within each test, both items and subtests are discarded unless they have a positive correlation with the test as a whole; and when a new test appears, it is simply vetoed unless it has a positive correlation with those tests already accepted. If this is what creates the consistency of performance across the diverse tests and items that *g* measures, the general factor is an artifact of how psychologists validate IQ tests. They rig the tests to produce *g* and then hail it as significant.

Jensen (undated, pp. 4—5) replies that when a variety of items selected only because they have some plausible claim to test mental abilities are given to representative samples of a population, there are many positive correlations, very few negative ones, and the positive correlations are always higher. As sample size increases, the negative correlations tend to disappear, suggesting that they are largely due to measurement error. He challenges anyone to take negatively correlated items and try to construct a mental test. Jensen (1982, p. 132) notes that Thurstone spent years trying to design tests that would not correlate, tests that would measure only verbal, numerical, or spatial factors, and that he failed: when administered as a collection, these tests so intercorrelate that a second-order *g* emerges which accounts for twice the variance of the special factors combined. Finally, Jensen (1980a, p. 230) isolates tests which have low *g* loadings, that is, correlate least with other mental tests. These turn out to be things like counting dots, making dots, crossing out designated letters or numbers, tests that hardly seem to be mental tests at all. They also have low external validity, that is, show little correlation with people's eventual academic achievement or socioeconomic status.

However, Block and Dworkin (1977, pp. 444—447) argue that the correlations with scholastic and occupational success have also been built into the tests. They note that Alfred Binet, the inventor of the first really useful mental test, used teachers' judgments of intelligence as a guide to selecting test items; and that since then, test after test has been discarded because it did not yield the proper correlations which means that extant tests are simply those which have jumped the correlational hurdles. They add that educational attainment is an important avenue towards occupational success, so rigging the tests in favour of the former automatically produced a correlation with the latter. In reply, it is mistaken to think that when a certain criterion has been used to screen test items, this provides a sufficient explanation of the correlation between the overall test score and that criterion. Binet did use teachers' assessments, that is, he observed the children described as 'bright' and selected his items from among the everyday tasks they could perform. But he was soon able to predict which children would fail in school better than the teachers themselves. As we have seen, present-day IQ tests do not merely reflect teachers' assessments but improve on them: they correct for irrelevant factors like docility, extroversion, and age-advantage.

Once a test isolates the mental processes that lie behind a correlation the magnitude of that correlation can deviate from the built-in value. Ravens is a good example of this: as a test of fluid *g*, it measures mainly mental ability only and thus has a lower than usual correlation with academic achievement; tests of crystalized *g* measure mental ability plus effort and thus have higher correlations.

Jensen (1979, pp. 83—93; 1980a, pp. 229—232 and 247—248; 1980b, pp. 365—366; 1982, p. 134) defends both *g* and IQ tests by way of a whole range of correlations that no-one has ever built into the tests. Moving from items and tests with high *g* loadings to those with low *g* loadings, the contrast is clear: a move from problem-solving to mechanical skills, from cognitive complexity to simplicity, from manipulation of materials to simple feed-back. It seems significant that items lose their *g* loading if not enough time is given for reflection, or if items become so complex subjects fall back on trial and error. Everyday life provides an example of a correlation with cognitive complexity: the task of making jelly-rolls is more *g* loaded than the simpler one of scrambling eggs. Learned tasks in general are more *g* loaded when they require conscious mental effort rather than mere memorization, when they are hierarchical (later learning is dependent on earlier learning), when they must be transferred out of the original learning context, when proofs rather than theorems are learned, and so forth.

Binet did not use intelligence assessments from other cultures to design his tests. Jensen grants that other cultures may value certain skills more than we do, for example, hunters may put a higher value on speed and motor coordination than on abstract problem-solving. However, the individuals they call intelligent tend to exhibit *g*. The Kalahari Bushmen of Africa call some of their tribe the 'clever ones' and these tend to score better than average on performance IQ tests. In fact, *g* has some application even across species. Human children can be given certain of the performance tests designed for animals and this reveals that *g*-loaded tasks put apes ahead of monkeys, monkeys ahead of dogs, dogs ahead of chickens. Gould (1981, p. 318) expresses horror: speaking as a paleontologist, he accuses Jensen of ranking all animal species, each of which possesses its own solution to its own environmental niche, according to human standards. Surely that is the whole point: human beings rank animals using a distinctively human concept of intelligence, the primitive concept found in everyday life, and these rankings correlate with *g*. Moreover, from time immemorial, human beings have associated low intelligence with certain genetic abnormalities and with inbreeding or the practice of mating with close relatives. There is a negative correlation between *g* and inbreeding, as measured by the offspring of cousin marriages in Japan, and carriers of the *recessive* gene for PKU (cretinism) average 10 IQ points below their noncarrier full siblings. Vogel (1980, p. 358) adds that IQ tests have proved sensitive to quite subtle effects from known genetic conditions, a spatial defect from Turner's syndrome and a slight verbal weakness from the recessive PKU gene.

In sum, Jensen believes that these correlations put the link between *g* and the primitive concept of intelligence beyond reasonable doubt. I believe he has earned the right to back *g* as a potentially useful scientific construct — and to hope that the evidence will vindicate his choice.

The Ontology of g

The case for the reality of *g* rests on four assertions: (1) That it describes a phenomenon, namely, the tendency towards consistent performance across mental tests; (2) That it measures an ability, that is, when *g* is operationalized it reveals that people have an advantage across mental tests because they possess a limited set of interrelated skills; (3) That it plays a causal role in the real world, that is, has explanatory power concerning the life-histories of individuals and groups; (4) That it has a physiological substratum, that is, correlates with certain elementary cognitive tasks and with evoked electrical potentials of the cerebral cortex.

The first and second assertions have already been established and the second adds much to the first, indeed, the fact that Ravens has operationalized *g* is crucial. Those who can solve perceptual analogy items possess *g* and this forecloses the possibility that so-called general ability is no more than an average of performance on functionally unrelated mental tasks. In sport, it is discouraging that no-one has been able to operationalize a general ability factor and it is no substitute to simply average someone's scores for archery and golf and pole-vaulting, any more than it makes sense to average marks for cooking, cleaning, and gardening. However, the first and second assertions have an important limitation, namely, they both apply only to the world of mental tests.

The case for a causal role means that for the first time, aside from intimations that the venture might prove worthwhile, we will be attempting to link *g* to the real world. On the personal as distinct from the group level, *g* is supposed to have a potent effect on academic achievement and a significant though lesser effect on socioeconomic status (SES), that is, on a person's ability to qualify for certain occupations and enjoy upward social mobility. But correlations cannot in themselves establish causal links between *g* and these variables. The claim has often been made that IQ tests are no more than academic achievement tests in disguise and, therefore, of course they correlate with academic achievement, just as academic achievement tests correlate with one another (Green, 1974). As for the correlations between IQ and SES, Block and Dworkin (1977, pp. 411—414, 431—432, and 453—454) argue that IQ tests may be little more than measures of social privilege and attitudes that go with privilege such as positive self-image and persistence. Therefore, good performance on IQ tests may be essentially an epiphenomenon, that is, pure effect with no causal role of its own. Someone from a privileged home would both do well on IQ tests and find an easy path to high SES; but privilege rather than IQ would be the real causal factor.

Jensen cites the work of Crano (1974; also Crano, Kenny, and Campbell, 1972), Gibson (1970), and Waller (1971) concerning the relationship between IQ and attainment. He stresses that early IQ predicts later academic achievement better than early academic achievement predicts later IQ. This is based on 5495 pupils who took both kinds of tests in Grade 4 (ages 9—10) and also took them two years later in Grade 6 (ages 11—12). Jensen (1980a, pp. 241—242) concludes that we are dealing with two distinct variables and that the causal line runs from individual differences in IQ towards individual differences in academic achievement, rather than in the reverse direction. This conclusion rests on an unstated assumption: that achievement differences should come into line with ability differences over time given that ability remains relatively stable. The unstated assumption makes sense if we also assume that children tend to perform above or below their ability at earlier ages, pushed at home or neglected, and that these discrepancies should disappear later thanks to formal education and a more uniform learning environment in general. These assumptions are at least plausible: the tendency of force-fed children to lose their achievement advantage as they progress in school is common.

The apparent effects of IQ were not uniform. As Brody and Brody (1980, p. 335) point out, when the sample was divided into mainly white suburban children and mainly black inner-city children, the predictive advantage of IQ was enhanced for the former but absent, even slightly outweighed by academic achievement, for the latter. When the sample was divided into high-IQ and low-IQ children, IQ had a predictive advantage for both but the advantage was much greater for the high-IQ subjects. I believe that these so-called disparate results fall into a coherent pattern. In a suburban middle-class school, mental ability claws its way towards a commensurate academic achievement. But when a child enters an inner-city school, a serious academic deficiency feeds on itself: with every year that passes, the child falls further and further behind. The tendency for initial differences in academic achievement to widen overwhelms mental ability as a causal factor. On the other hand, the higher the level of mental ability, the more it has the power to assert itself. All of the results are at least consistent with the hypothesis that mental ability, as measured by IQ, is a real-world causal force.

Studies of social mobility show that IQ has a differential effect on the life-histories of individuals even when social privilege is held constant.

In England, when siblings reared in the same family differ significantly in IQ, those above the family average tend to move up the SES scale and those below to move down. In America, a son who has an IQ 8 or more points above or below his father's (taken at the same age) tends to attain higher or lower social status accordingly. When the difference is as much as 23 points, the tendency is overwhelming. The average IQ difference between parent and child, and also between siblings, is 12 points and more than 50 per cent of children move into a different SES category as adults.

Jensen (undated, pp. 36—40) makes an important supplementary point. Although more evidence is needed, he gives results from three samples each composed of subjects who took a collection of mental tests and were numerous enough to include many siblings. The performance of sibling and co-sibling is the closest we can get to simulating a classless society, given that class differences operate primarily between families. Jensen did a factor analysis of the tendency of sibling and co-sibling towards consistent performance across mental tests, one tending to do better the other worse. The *g* that emerged was virtually identical to that extracted from the performance of members of different families who, of course, represent our present society with all its class divisions. Since *g* remains unaltered when no classes exist, it cannot be primarily a measure of social privilege. Jensen concludes that it appears to measure an ability that ranks people within as well as between classes. I agree, although this does not commit me to the ultimate conclusion that the ability measured can be identified with the primitive concept of intelligence.

Ever since Spearman invented factor analysis, the adherents of *g* have yearned for a physiological substratum. Gould (1981, p. 310) cites the fact that "no concrete tie has ever been confirmed between any neurological object and a factor axis" as crucial. Jensen has taken up this challenge on two fronts.

First, Jensen (1980a, pp. 686—692) and others have discovered correlations between the performance of subjects on Ravens, used as a marker test to measure how much subjects differ for fluid *g*, and their reaction times (RTs). Simple reaction time measures how long it takes to remove the index finger from a 'home' button so as to press another button when a light adjacent to the latter goes on. Choice reaction time measures the time taken to release the home button when the subject is confronted with a choice, say sees four digits on a screen, then is shown one digit, and must decide whether the latter was one of the former by pressing a 'yes' or 'no' button. The highest correlations are obtained between *g* and choice reaction times and usually range from 0.35 to 0.45; these are of course rather low accounting for only 12% to 20% of variance, that is, for only a small percentage of individual differences in *g* (variance accounted for = correlation squared, for example, $0.35 \times 0.35 = 0.12$).

Second, as an even closer approach to a physiological substratum, Jensen (1980a, pp. 707—710; undated, pp. 31—33) cites correlations between *g* and various measurements of the brain's electrical potential. Two or more small electrodes are gummed to the subject's scalp and a flash of light or a 'beep' is presented at random intervals every few seconds over a session of about 10 minutes. The experimenter can view the subjects' 'brain waves' on an oscilloscope but the real readings are done by computer. The computer scans the brain's electrical responses to the stimuli over the entire session arriving at an average evoked potential (AEP) which represents the brain's characteristic response. The principal readings are AEP latency, the time between the stimulus and the brain's response, AEP amplitude, the height of the graphically recorded waves, AEP complexity, a measure of the shape of the waveform, and AEP habituation, the difference between wave height (amplitude) during a first session of exposure to stimuli and a second. Until recently, correlations were very low but Eysenck and Barnett (1985) report a value of 0.60 between Wechsler IQ and AEP complexity, Scafer (1985) 0.73 between Wechsler IQ and AEP habituation, Wechsler tests being marker tests for a mix of fluid and crystallized *g*.

While granting that this research is at an exploratory stage, Jensen (1980a, pp. 700—704 and 707—708; undated, pp. 33—34 and 41) believes it has great theoretical importance. He believes the RT correlations alone completely refute the notion that individual differences in *g* are largely the result of individual differences in learned strategies. Students may learn strategies that give them 'short-cuts' to solving mathematical problems on achievement tests, but when they excel on Ravens, they do so because of the superior speed and efficiency with which they can execute basic cognitive operations, such as how quickly things held in short-term memory can be scanned and retrieved. He predicts that an item's *g* loading will turn out to be a function of the number or importance of the neural processes that are involved in the item's solution. The AEP correlations also dispel the notion that *g* is no more than skills people learn at home or in the larger culture for, after all, the brain's electrical potential is even further from cultural influence than reaction times. He believes the AEP research may prove to be the 'Holy Grail' that links *g* with the brain's neural efficiency; eventually correlations between *g* and neural structure may be found all the way down to a level just short of the molecules or even atoms that compose the brain.

Later we will have reason to refer back to Jensen's contention that *g* is an ability dependent on the neural substratum of mental activity and not much influenced by learned strategies or cultural factors. For now, I have tried to give a fair summary of his views about the ontological status of *g*. This is important because Gould (1981, pp. 24, 151, 159—160, 239, and 317—320) launches a fierce attack on Jensen on this very point. Jensen above all has committed the sin of reification: he takes the diverse set of capabilities called intelligence and labels them with an abstraction called *g*; having reduced them to one concept, he then converts that concept into a thing located in the brain. Well, Jensen's views are more complicated than that. If he can show that *g* measures a tendency to do well on a variety of mental tasks, that it can be operationalized as an ability, that it has true causal potency, and that it has a physiological substratum, he can make the appropriate claims about its reality or ontological status. If he cannot show these things, the claims must be withdrawn. Perhaps the sin of reification is making such claims without sufficient evidence; if so, we would do well to drop the word and focus on the evidence.

Sometimes Gould seems to be simply saying that intelligence is more than academic intelligence, that even the latter includes diverse abilities, and that we overlook this diversity at our peril, for example, how could anyone give a sensible diagnosis of a case of mental retardation just by saying a person is deficient in *g*? These points have already been conceded. But once again, the fact we cannot measure all forms of intelligence should not prevent us from measuring academic intelligence, the fact that academic intelligence comprises many abilities does not mean they are all as scientifically interesting as *g*, and the fact that *g* cannot tell us everything does not mean it cannot tell us anything.

THE FALSIFICATION OF *g*

Arguing both that the merits of a theory have often been overlooked and that it must be revised and transcended encourages a certain amount of schizophrenia. Thanks to the above account, the Spearman—Jensen theory may appear to have enjoyed success after success and stand at the apex of its scientific plausibility. It is time to set the record straight: over the last 40 years, psychologists have discovered something that the theory can neither explain nor accommodate. I refer to the phenomenon of massive IQ gains over time, that is, the tendency for each generation to have a higher average IQ than its predecessor. Until recently, the theory progressed unaffected by this phenomenon, but that was because reports of massive IQ gains were either little known or largely ignored.

The interaction between the Spearman—Jensen theory and the phenomenon of IQ gains over time is a classic example of the degeneration of a theory in the face of evidence. The following is borrowed from Lakatos with a few changes in terminology: (1) A theory generates a series of *original* hypotheses which flow naturally from the theory; (2) The theory itself cannot be falsified but its hypotheses can; (3) If that happens, defenders of the theory put forward *protective* hypotheses — these are designed to defend the theory but may well make interesting predictions and possess scientific respectibility; (4) If these in turn are falsified, defenders of the theory may begin to put forward *ad hoc* hypotheses — these too defend the theory but at the price of doing nothing else and are scientifically bankrupt; (5) When it becomes clear that nothing better than *ad hoc* hypotheses are forthcoming, it is time for the theory to be modified or replaced by another.

The Theory and Its Hypotheses

The Spearman—Jensen theory has three essential parts: IQ tests measure a mental ability called *g*; *g* bears a close resemblance to the primitive concept of intelligence; intelligence is a mental ability with a potent causal role, that is, it is productive of achievements of the kind usually associated with academic ability. How closely *g* matches popular notions of intelligence is not important. But if it did not refer to a mental ability other than learning or memory that played the above causal role, the theory would not deserve to be called a theory of intelligence at all. The theory logically entails the following hypothesis: if the people of a modern industrialized nation make massive gains in *g* over time, they will make corresponding gains in terms of creative achievement, scientific and mathematical discovery, and technological progress.

This hypothesis implies a comparison between groups, that is, an ability comparison between the present generation of a given nation and past generations. Up to now, our account of the theory has focused on its explanatory power concerning the life-histories of individuals, but its proponents have always assumed that group differences in IQ or *g* are highly significant. Some group differences are quite large, ranging from 10 to 15 or even 20 points. These can be best appreciated if we take into account that IQ scores have a normal distribution with a standard deviation (*SD*) of 15. This means that if a particular group has a mean IQ 15 points below another, only 16 per cent of its members will exceed the average of the superior group and, in addition, it will be on the wrong end of a 17 to one ratio at high IQ levels, that is, 130 and above. A deficit of 20 points means that only 9 per cent of the inferior group will exceed the superior group's average and now, the ratio will run 57 to one against it at high IQ levels.

When Lynn (1982) discovered that Japan was 11 points above America on the performance half of the Wechsler, he suggested that this might have been a significant factor in Japan's outstandingly high rate of economic growth since World War II. Nichols (1987) focuses on the fact American blacks are 15 IQ points below American whites and argues that, thanks to that, nothing can be done about how unfavourably blacks compare to whites in terms of real-world achievement, indices such as income, occupational status, symptoms of family demoralization, and crime statistics. Eysenck (1985) sums up his survey of racial differences as follows: Mongoloid peoples have the highest mean IQ, particularly the Chinese and Japanese, and the scores then decline through Northern Europeans and their descendants, through Southern Europeans and Indians, down to Malays and Negroid groups at the bottom. He sees a close correlation between mean IQ, socio-economic status, and level of cultural achievement. Harlow and Harlow (1962, p. 34) generalize: human beings collectively have little more than the minimum intelligence needed for social progress and a mean IQ significantly below that level would mean that there could be no civilization as we know it.

Jensen (1980a, 1981, 1985) has made a detailed comparison of American blacks and American whites. IQ has the same moderate success in predicting academic achievement, as measured by standardized tests and university grades, for blacks as it does for whites. There is a slight tendency for blacks to underachieve as compared to whites with the same IQ. The primary factor that differentiates black and white test performance is a racial difference for *g* and the underrepresentation of blacks in elite occupations makes sense in terms of *g* differences between the people who normally staff various occupations. Black and white have similar hierarchies of item difficulty on IQ tests which means the same items are difficult or easy for both. Insofar as black children of a given age have a different item hierarchy from whites of the same age, it mimics a maturity difference, that is, the performance of black 13-years olds is indistinguishable from that of white 11-year olds. The tallies with the hypothesis that black—white *g* differences account for black—white test score differences in that fluid *g* increases with age up to about 18 to 20. Jensen concludes that the 15-point black IQ deficit represents a real 'mental maturity' or ability difference between the races, rather than the effects of tests culturally biased against blacks.

Jensen's analysis shows that in many respects, American blacks perform as if they were a group of less able whites, a group selected out of the white population on the basis of below-average IQ. However, blacks differ from that pattern in other respects. When black children are compared to their own parents, their IQs regress to their own population mean rather than to the white mean, reflecting the fact that they constitute a separate breeding group with its own environment. They have much lower incomes and socioeconomic status than would a group of whites that match them for IQ (Flynn, 1987a) and they self-select in terms of occupational aspirations by way of ability comparisons with one another rather than with the larger white community (Jensen, 1980a, p. 101).

When between-group hypotheses compare Japanese and Americans or Northern Europeans and Malays, they may not seem to be a fair test of the Spearman—Jensen theory. When groups speak different languages, practice radically different customs, have different histories and incentive-systems, stand on opposite sides of the industrial revolution, and so forth, this raises questions about the ability of IQ tests to bridge cultural distance. The proponents of the theory argue that the right tests correctly administered can answer such questions, but they may be ignoring common-sense limitations on its explanatory power and a theory should not be measured against the audacity of its proponents. On the other hand, between-group hypotheses about successive generations within nations, ones long part of the technologically developed world, should constitute an ideal test of the theory: common language, common history, continuity of custom, tests long introduced and easily administered. If the theory is to generate no hypotheses about such groups, it is difficult to see that it can lay claim to any between-group explanatory power.

As we have seen, massive IQ gains from one generation to another enormously increase the proportion of the population at high IQ levels. Jensen (1980a, pp. 111—114) makes clear what real-world behaviour we have a right to expect from those at high IQ levels: above 130 they find school easy and can succeed at virtually any occupation; above 140 their adult achievements are so extraordinary they fill the pages of *American Men of Science* and *Who's Who*; above 150 they amaze their teachers with their precocity and begin to duplicate the life histories of famous geniuses who made creative contributions to our civilization. Jensen asserts that the quality of a society's culture is highly determined by the fraction of its population that is highly endowed and that they are the main source of philosophical insights, mathematical and scientific discoveries, practical inventions, masterpieces of literature and art, and so forth. Clearly massive IQ gains should bring a cultural renaissance too obvious to be overlooked.

The hypothesis that massive IQ gains will mean corresponding gains in terms of real-world achievement is implied by the Spearman—Jensen theory, but the theory has an accompanying expectation, namely, that generational IQ gains will be small. Large intelligence differences between races and between nations may seem plausible, at least to some, but huge intelligence differences between generations seem absurd: we live in intimate daily contact with our children and they do not appear that much more brilliant than ourselves and we do not remember our parents as being intellectually limited. Therefore it is not surprising that for 40 years studies reporting massive generational IQ gains were largely neglected, although admittedly many of them simply did not circulate across national boundaries. Here is a partial list of studies, each followed by the reason for its neglect: Tuddenham (1948) — U.S. Army tests said to be more akin to achievement than intelligence tests; Elley (1969) — New Zealand results and thus ignored in accord with international custom; Bouvier (1969) — Belgian Army pamphlet published in French only; Thorndike (1973) — doubts about the American Binet samples; Girod and Allaume (1976) — published in French only; Mehlhorn and Mehlhorn (1981) — East German results published in German only; Rist (1982) — Norwegian Army pamphlet published in Norwegian only; Lynn (1982) — emphasized comparison between Japan and America rather than Japanese gains; Flynn (1984) — Wechsler test gains in America seemed offset by Scholastic Aptitude Test (SAT) losses; Leeuw and Meester (1984) — published in Dutch only; Schallberger (1985) — technical report in German only.

However, massive IQ gains have become so universal throughout the technologically developed world and so well documented that they can no longer be dismissed. It was something of a turning point when Lynn, long skeptical about the radical malleability of IQ, accepted the phenomenon as real and actually began to add to the evidence (Lynn and Hampson, 1986). Flynn (1987b) has analyzed post-1950 data from 14 nations: the present generation has a mean IQ 5 to 25 points above the last; the advantage varies from nation to nation with a median of 15 points or a full standard deviation. Some of the most impressive gains have been in Western Europe, for example the Netherlands gained 20 points from 1952 to 1982 and France some 20 to 25 points from 1949 to 1974 with 20 points being acceptable as a conservative estimate. Therefore, the potentially creative elite of both of these nations multiplied over that period by about 57 times.

Flynn (1987b) verified that neither the Netherlands nor France had any perception that they were in the midst of a cultural renaissance. There is not a single reference to a dramatic increase in genius or mathematical and scientific discovery in the present generation. No one has remarked on the superiority of contemporary schoolchildren. The number of patents granted for inventions has diminished in both nations. Jensen (1987) concurs that if the real-world achievements we associate with intelligence had escalated in accord with massive IQ gains, the results would be too obvious to be missed: the aging survivors of the last generation would be perceived by almost everyone as border-line mentally retarded; Dutch university professors would be amazed at the prevalence of genius in their classes.

The facts are these: when IQ tests rank people at a given place and time the results make sense in terms of relative intelligence; when IQ tests rank generations over time they give nonsense results. These facts imply that IQ tests do not measure intelligence but rather a correlate with a weak causal link to intelligence. Imagine we could not directly measure the population of cities but had to take aerial photographs which gave a pretty good estimate of area. In 1952, ranking the major cities of New Zealand by area correlated almost perfectly with ranking them by population and in 1982, the same was true. But if anyone found that the area of cities had doubled between 1952 and 1982, they would go far astray by assuming that population had doubled. The causal link between population and its correlate is too weak, thanks to other factors that intervene such as central city decay, affluent creation of suburbs, more private transport, all of which can expand area without the help of increased population.

The implications for IQ tests as measures of intelligence is serious: their explanatory power appears to be a function of cultural homogeneity. They are at their best when accounting for the differential real-world achievements of siblings raised within the same family; they do reasonably well with Dutch or French or Americans of the same generation sharing the same formative environment; they fail completely when they attempt to bridge the cultural distance that separates generations in modern industrial societies. This last means that all between-group comparisons which attempt to bridge even greater cultural distance are suspect. I refer to Lynn's use of IQ differences to explain differential real-world achievements by the Japanese and American economies and all other cross-national and cross-racial comparisons.

It may be said that to reject Jensen's comparison of American blacks and American whites is question-begging. If blacks are so assimilated that they mimic a subgroup of the white population, the question of cultural distance does not arise and the racial IQ difference could be identified with an intelligence difference. In response, some of Jensen's evidence is less impressive after analysis of the between-generations data. Recall that the fact black 13-year olds mimic the test performance of white 11-year olds was used to suggest a real ability difference between the races. W. B. Dockrell (personal communication, 7 August 1985) has supplied data on Scottish IQ gains between 1965 and 1982. These are greatest on those Wechsler subtests that measure reasoning ability such as similarities and comprehension. The performance of yesterday's 13-year olds on these two subtests mimics the performance of the 10 or 11-year olds of today. So the generations also appear to be separated by a real ability difference and yet, unless evidence of extraordinary achievement is forthcoming, that ability difference cannot be identified with an intelligence difference. At the very least, when we have new tests that are better measures of intelligence, blacks will get a second chance to improve their performance.

The between-generations data also solves a problem I have always found disturbing. Elsewhere I have attempted to show that the black IQ deficit is probably due to environmental rather than genetic inferiority (Flynn, 1980; 1987c). It hardly seemed plausible that environmental differences between the races could cause *g* differences and nothing else; that is, no differences such as blacks being less familiar with certain test items than whites. And yet, black and white item hierarchies differed in terms of *g* loading and little else — that was why older blacks mimicked the performance of younger whites. The between-generations data proves that all of this is possible: the generations are separated almost entirely by environmental differences; and yet, the older Scottish children of yesterday mimic the younger Scottish children of today.

It is important to avoid confusion. In order to shed light on this particular black-white problem, generational IQ differences must reflect *g* differences only, at least in one nation. But generational IQ differences threaten the Spearman—Jensen theory just so long as they signal massive *g* gains unaccompanied by achievement gains. Additional gains or losses caused by enhanced or diminished familiarity with items can provide an unwelcome complication but they in no way reduce the threat.

Protective and Ad Hoc Hypotheses

The proponents of a theory under threat can either question the evidence or surround the theory with a belt of protective hypotheses or both. The proponents of the Spearman—Jensen theory of intelligence have done both.

From 1982 to 1984, the English-speaking world was aware primarily of evidence for massive IQ gains in Japan and America. Wechsler and Stanford—Binet standardization samples, selected over a period of a generation or more, had performed better and better on a variety of Wechsler and Binet tests. The samples were stratified rather than random, numbered from about one to three thousand subjects, and in most cases test content varied somewhat because of revision or the age-group for which the test was intended. Jensen (1980a, p. 570; personal communication, 3 February 1983) stressed the possibility of sample bias: he stipulated that samples should be extremely large, say comprehensive testing of draft registrants, and the tests identical. These conditions have now been met: Dutch, Norwegian, and Belgian IQ gains are based on military testing of virtually the entire population of young adult males and the tests were unaltered.

Turning to protective hypotheses, these are entirely respectable just so long as they suggest testable predictions. When Newton's gravitational theory failed to explain the orbit of Mercury, the hypothesis offered was that it was being influenced by an undiscovered planet closer to the Sun. The orbit required for the undiscovered planet was calculated, it was tentatively named Vulcan, and the French Academy awarded the Legion of Honour to M. Lescarbault for its discovery. The sighting was wishful-thinking and Newton's theory had to be transcended, but the hypothesis was potentially fruitful. The planet Neptune had been discovered in precisely this way, that is, as a posited influence on the orbit of Uranus.

When the Spearman—Jensen theory failed to explain the phenomenon of massive IQ gains, there were three obvious targets for protective hypotheses: the subjects who took the tests, the tests themselves, and the character of the enhanced performance.

Hypothesis: Massive IQ gains represent early maturation rather than gains at full maturity (Jensen, 1980a, p. 570; personal communication, 12 January 1983). If generational gains were large among young children, small among older children, and non-existent at full maturity, there would be no reason to expect enhanced adult achievement and lack of such would not weaken the link between IQ or *g* and intelligence. Full maturity refers to the age of peak raw-score performance on a particular test. It is not easy to establish such an age for tests of crystalized *g* because they emphasize acquired knowledge and additional knowledge can be acquired into old age. However, performance on tests of fluid *g* like Ravens maintains a stable maximum from ages 18 to 25, then begins a gradual decline which accelerates in old age. *Prediction*: IQ gains will decline with age among school children and will be absent by ages 18 or 19 on tests of fluid *g*.

Hypothesis: Massive IQ gains represent learned item gains rather than *g* gains (H. J. Eysenck, personal communication, 14 December

1982; Jensen, 1980a, pp. 569 and 635; Jensen, personal communication, 12 January 1983). Tests of crystalized *g* emphasize acquired knowledge and therefore, really measure mental ability or *g* only when all subjects have had an equal opportunity to acquire the knowledge in question. This situation may obtain for all or most members of a particular generation but not from one generation to another: over 30 years, general educational advances may give the current generation a large advantage over the last on learned content items. Take Plato or Aristotle or Archimedes: they would do badly on a modern general information subtest or an arithmetic subtest using modern notation. However, they should do very well indeed on tests of fluid *g* like Ravens. Once they had become familiar with the directions and multiple choice format, they would need to demonstrate only decontextualized problem-solving ability and their performance would be a true measure of *g*.

As the elaboration of this hypothesis makes clear, it protects the Spearman—Jensen theory by arguing that massive IQ gains may represent only a larger repertoire of learned items and not massive *g* gains. Without massive *g* gains, we would expect no escalation of intelligence or real-world creativity hence lack of such does not count against the theory. There are difficulties with this hypothesis: on the one hand, we are told that the knowledge IQ tests require is so elementary and universally accessible that the tests are fair measures of *g* differences between subjects; on the other hand, we are told that this knowledge was so poorly acquired by one generation that huge improvements were made by the next, so that the tests are not fair measures of *g* differences between generations. At any rate, this hypothesis clearly tells us what to expect. *Prediction*: IQ gains will be found primarily on tests of crystalized *g* and be small or absent on tests of fluid *g* such as Ravens.

Hypothesis: Massive IQ gains represent enhanced test sophistication rather than *g* gains (H. J. Eysenck, personal communication, 14 December 1982; J. C. Loehlin, personal communication, 3 January 1983; J. Ray, personal communication, 7 August 1986). Once again, the theory is protected by breaking the link between IQ and *g*. Test sophistication refers to the fact that people who have never taken formal tests or a particular kind of test are unfamiliar with format and strategy. When first confronted with multiple choice items, they may go through the series of suggested answers in the order presented, rather than scanning them to see if there is an obviously correct choice.

Therefore, they are at a disadvantage compared to more sophisticated subjects in a way that has nothing to do with mental ability or *g*. Fortunately, test sophistication has been much studied and the gains involved carefully estimated. Jensen (1980a, pp. 590—591) concludes that even working with totally naive subjects, repeated testing with parallel forms of IQ tests gives gains that total only 5 or 6 points; moreover, increments after the first exposure to testing rapidly diminish and approach a threshold or limit. *Prediction*: IQ gains will exhibit a declining rate of gain and approach a limit of 5 or 6 points.

Flynn (1987b) shows that all of the predictions engendered by these protective hypotheses have been overwhelmingly falsified. Over the last 30 years, there have been massive gains on tests of fluid *g* in the Netherlands, Norway, France, and Australia, gains ranging from 12 to 25 points. Belgium is also impressive with 7 points in only 9 years. Lest it be thought that Ravens is the only test of fluid *g* that shows gains, the Australian gains are on Jenkins, the Belgian Shapes Test matches their Ravens gains, and recently Lynn, Hampson, and Mullineux (undated) have shown that British gains on the Cattell are actually greater than British Ravens gains. Gains on tests of crystalized *g* are sizable but lag behind tests of fluid *g*. In other words, fluid *g* gains are actually greater than the typical IQ tests would convey: the acquired knowledge content of the typical test actually disguises the magnitude of *g* gains rather than explaining them away! The Netherlands, Norway, France, and Belgium all give us subjects aged 18 or 19 which means they are fully mature in terms of fluid *g*. Five nations give data for school children of various ages and four of these show no tendency for gains to decline with age.

As the above shows, generational IQ gains do not follow the pattern of test sophistication gains. They are usually far greater than 5 or 6 points and while trends over time vary, there are nations like the Netherlands where gains have continually accelerated over 30 years rather than declining as they approach a limit. It may be suggested that perhaps some new form of test sophistication has appeared which gives average gains three or four times as great as anything we know and which feeds on itself. This is purely a verbal strategem. Test sophistication proper refers to the limited gains arising out of familiarization with test setting and format. To turn it into something radically different is merely to give cases in which people have more ability of the sort IQ tests measure a new name, perhaps 'super test sophistication'. This is like attributing the ever better performance of athletes to some new and incredibly potent form of 'competition sophistication'. The actual experience of competition produces limited gains that are well known and there is no point in applying the term to a quite different phenomenon.

In a recent publication, Jensen (1987, pp. 379—381) rehearses some of the protective hypotheses, emphasizes that IQ gains have not meant intelligence or real-world achievement gains, and concludes as follows: "It seems much more plausible that the reported test score increase of twenty points does not reflect a corresponding change in *g* or its real-life correlates, but is rather the result of some artifact not yet identified." Given the context, this assertion could mean several things. Jensen was reacting primarily to the Dutch data only and perhaps all he means is that we should wait until all the evidence is in concerning the protective hypotheses stated thus far. All well and good, but the evidence is now in and the protective hypotheses have been falsified. Perhaps the reference to an "artifact not yet identified" is crucial. This would be tantamount to an expression of hope that some new as yet unformulated protective hypothesis will come along that will not be falsified, that a 'factor X' will be discovered that successfully severs the link between massive Ravens gains and massive *g* gains.

The proponents of the Spearman—Jensen theory can legitimately ask for a breathing space to rethink their position. However, such a period should not be too prolonged. Jensen himself (1973a, pp. 135—139 and 188—189; 1973b, pp. 351 and 413—414) has ridiculed those who have nothing better to offer than an unknown 'factor X' that yields no testable predictions, for example environmentalists at a loss for an explanation of racial IQ differences. He also points out the extraordinary difficulty of finding a factor that will explain away between-group differences and not also explain away within-group differences. For example, he believes that any plausible factor that differentiates blacks from whites would differentiate blacks from one another. Now he is in precisely the dilemma he describes: finding an artifact that will work between generations but have no effect within generations. The Spearman—Jensen theory cannot afford an artifact that severs the link between Ravens and *g* for individual differences within a generation, as well as severing the link between generations. Ravens is the marker test for fluid *g* in all their physiological research; IQ tests would be as suspect in ranking individuals for intelligence as they are in ranking groups. And yet, if 'super test sophistication' or 'acquired knowledge' varies from one generation to the next, how likely is it that there is no significant variance for the individuals who comprise a particular generation?

Jensen's assertion that generational IQ gains cannot be plausibly construed as *g* gains is susceptible to a third interpretation. At times, he refers to the well-established correlates between *g* and real-world achievements and uses the phrase "intelligence or *g*" as if the two were interchangeable. Perhaps Jensen would reject the following interpretation, but I will proceed because I have no doubt that it will tempt many of the proponents of the Spearman—Jensen theory.

The theory has had many successes in evidencing links between *g* and the real-world achievements we associate with intelligence. Nevertheless, these cannot justify contending either that its successes cancel out its failures or that when the appropriate real-world achievements are absent then *g* simply must be absent as well. These confusions can be dispelled by keeping in mind exactly what *g* is. It begins as no more than an ability operationalized by Ravens which allows people to do well across IQ tests. Every bit of its real-world explanatory status must be won through evidence and no bit carries over to another bit. Between-group differences are supposed to be a major area of explanation: Jensen (1980, p. 636) says Ravens can measure fluid *g* for groups of people of remotely different cultures. Its between-generations failure can no more be cancelled out by its successes than Newton's failure with Mercury could be cancelled out by successes with the other eight planets. Indeed, it was the gravitational theory's great success that made Mercury such a scandal.

As for positing that *g* must be absent whenever the real-world achievements of intelligence are absent, this would be an *ad hoc* hypothesis of the worst sort: an undefended boundary hypothesis. The construct *g* cannot be deemed present or absent for the convenience of the theory. Saying *g* simply must not apply to between-generations IQ gains is no better than saying gravitation simply must be different when you get to the planet closest to the sun. This hypothesis generates no prediction save the embarrassing one that under certain conditions, hitherto thought similar to normal conditions, the theory will not work. However, it may be said that analogies with physics are all very well, but is it really unreasonable to posit the absence of *g*? After all, we have posited the absence of intelligence gains when real-world achievement gains were missing. Why not posit the absence of *g* for the very same reason?

The analogy looks plausible only because, for simplicity's sake, I have spoken of the primitive concept of intelligence as if it referred to something known to exist which can be here or there. As the last section of this paper will show, that is not at all its ontological status. It is really like an invitation issued to various scientific constructs, asking them if they wish to play a certain explanatory role, the role of a mental ability that causes certain real-world achievements. When they use Ravens, the proponents of the Spearman–Jensen theory accept the invitation: they use Ravens as the marker test for fluid *g* to give *g* a chance to play the explanatory role. When Ravens is there, that is *prima facie* evidence that *g* is there and if the appropriate real-world achievements are absent, that is *prima facie* evidence that *g* has failed. When an actor reads for a part and fails to perform properly, it is absurd to cite his or her failure as evidence the actor did not show up for the audition. You can of course claim it was someone else in disguise: that is what the theory's protective hypotheses tried to do — to say that learned items or test sophistication showed up disguised as *g*.

However, they failed and for now the limitation applied to IQ tests must be applied to *g*: it does not measure intelligence but rather a correlate with a weak causal link to intelligence; or put differently, the ability of *g* to play the role of intelligence is a function of cultural homogeneity.

Future of the Theory

Assuming that no protective hypothesis comes to the aid of the Spearman–Jensen theory, it must be revised or abandoned or transcended. However, no reasonably successful theory is abandoned or transcended until a better theory comes along, and for the time being our attitude to the Spearman–Jensen theory should be as follows: use IQ tests within clearly homogenous cultural settings; abandon group comparisons across cultural distance; use IQ tests for clinical purposes where clinical psychologists have found them of diagnostic value; be wary of taking their cutting lines for mental retardation at all seriously (Flynn, 1985). Their use in schools and as university entrance examinations cannot be discussed in a sentence or two. However, excellent performance on academic achievement tests should never be discounted because of lower IQ scores; academic achievement validates IQ tests and not the reverse. The kind of criterion sampling already described should replace IQ tests in selecting people for virtually all jobs and for armed forces training programmes as well.

Until a better correlate of intelligence than *g* is derived, the search for correlations with physiological data can continue but in a very tentative spirit. The weaknesses of *g* could be either inflating or deflating the physiological correlations and their existence certainly does *not* show that *g* is some kind of rock impervious to the usual cultural influences. Jensen is simply wrong on this point. Huge *g* gains from one generation to another show that it is highly sensitive to environmental factors and some of these may be cultural factors such as learned strategies of problem-solving picked up at school, or at home, or elsewhere.

As for changing the theory, the least radical change would be revision by way of better IQ tests. The present construct of *g* has failed and if you want a better *g*, you must improve the tests that engender *g*. I have no special expertise here, but I find some of the suggestions of Sternberg (1985) exciting. Clearly one criterion of a better test is that aside from matching the explanatory power of current tests, it must not give huge *g* differences where no real-world achievement differences exist. The prospect of a fundamental rethinking of mental tests should be regarded as exciting rather than a cause for gloom. Who knows what better tests and a better *g* might bring: a real understanding of group differences; higher physiological correlations? Who knows how many lines of research have been crippled by the defects of current *g*?

On the other hand, the failures of the Spearman—Jensen theory could necessitate a conceptual revolution: Einstein did not merely alter Newton's predictions in selected cases; he revolutionized the foundations of gravitation theory and our concepts of space, time, and light. If plugging a better *g* into the Spearman—Jensen theory does not work, its whole conceptual system may have to be replaced on a theoretical level. When a theory is replaced, this may take a more or less radical form, that is, it may either be abandoned or transcended. Like Velikovsky, it may be completely discredited because for every fact or supposed fact it explains, it generates several falsified hypotheses. Or like Newton, it may be transcended by a new theory: the conditions under which it works may be explained, its failures explained, and its predictions displaced only in certain cases where the stipulated conditions do not hold. I very much suspect that the Spearman—Jensen theory will not be abandoned but will find a place within a new theoretical structure.

The conceptual revolution that may be necessary can best be envisaged by positing what might at first appear to be a mere protective hypothesis: that massive *g* gains did bring massive intelligence gains but that the real-world achievements associated with intelligence were suppressed by other trends. Take the prediction the generational data falsifies: if the people of a modern industrialized nation make massive *g* gains over time, they will make corresponding gains in terms of creative achievement, scientific and mathematical discovery, and technological progress. This assumes, at least in the cultural context named, that intelligence is so incredibly potent a factor that dramatic effects should be visible no matter what else is going on. Perhaps academic intelligence has been diverted away from academic pursuits, perhaps factors like motivation have collapsed, perhaps educational institutions have so lost their way as to squander enhanced intelligence, perhaps a new character type means people do not express enhanced intelligence in the extroverted wit and sparkle which would make its presence evident. In other words, this kind of protective hypothesis would force us to face up to something everyone knows but all of us dislike because it complicates our task: the theory of intelligence can make no real progress until scientific constructs of intelligence are put in the context of larger theories of personal and social development.

The Spearman—Jensen theory has a tendency to treat *g* as if its explanatory potential could be assessed in virtual isolation from other factors. A decade ago Block and Dworkin (1977, pp. 416—419) demanded a more comprehensive theory to explain why IQ tests sometimes work; this demand can no longer be ignored now that we need to know why IQ tests sometimes fail. It is instructive to note how economics has reacted when its theories have faced similar failures.

For example, Keynes (1936) developed a general economic theory in which the disposable income available to households was the principal determinant of consumer demand. Within-generation data suggested that the percentage of income spent on consumption decreased as disposable income increased, that is, within each generation, the affluent logically enough spend a smaller proportion of their income on consumption than the poor. This led to a between-generations prediction that as people in general became more affluent, there would be a disastrous fall-off of consumer demand. In fact, the next higher-income generation spent the same percentage on consumption as the previous lower-income generation. Between-generations income differences did not produce the same effects as within-generation income differences. The mistake had been to assess the explanatory power of income in virtual isolation and once it was put in the context of a larger complex of variables, the failure was explained. A psychological variable was crucial: when people over time begin to believe that an increase in their incomes is permanent, they begin to spend a higher proportion on consumption and the average percentage is no less than when incomes were lower. The moral, of course, it that psychology may explain its between-generation failures when intelligence is related to other factors in a larger theoretical framework.

My own guess is that the paradoxes of massive IQ gains over time will be solved neither by better tests and a better *g* alone, nor by a more comprehensive theory of personal and social development alone. Social trends and a new character type may well have suppressed some of the real-world achievements associated with the primitive concept of intelligence, but it just does not seem plausible that the effects of such huge intelligence gains could be completely nullified. Therefore, tests of fluid *g* must be at fault when they suggest that the intelligence gains have been so large. Probably the Spearman—Jensen theory will have to be both revised and transcended.

The Primitive Concept of Intelligence

We end as we began with the primitive concept of intelligence. This concept should be construed not as something to be located or measured, but rather as a piece of advice to psychologists about research strategy. It says in effect: when you formulate a theory to explain the life-histories of individuals or groups, your theory will lose explanatory power unless it includes a mental ability or abilities distinct from memory and learning.

When Kepler realized that the orbits of the planets deviated from 'natural' or circular motion, he seized upon the primitive concept of celestial influence. He was a Pythagorean Sun-worshipper and the Sun was so big and so close it seemed that it just must have some effect on the planets that raced around it. The primitive concept of celestial influence said in effect: when you formulate a theory to explain the motions of the planets, your theory will lose explanatory power unless it includes the notion of influence by heavenly bodies that are large and proximate. The primitive concept had no specificity or explanatory power of its own. It was up to astronomers to supply that by putting forward theories and scientific constructs which would yield predictions that matched the phenomena in question. There have been a variety of such: Kepler speculated that the Sun acted as a magnet; Descartes that it rotated setting up a vortex or whirlpool that swung the planets around; Newton that its mass exerted a gravitational pull inversely as the distance squared; Einstein that its mass influences the shape of the space through which the planets move.

The latter constructs have had great explanatory power and have proved that the advice of the primitive concept of celestial influence was sound: our theories would be impotent if they omitted the facts of the Sun's size and proximity. But it would have been absurd to attempt to sharpen and measure the primitive concept. The thing to do was take its advice and find a scientific construct that had explanatory power within the context of a comprehensive theory that interrelated all the factors affecting celestial motion. The advice might have been bad, in which case all such theories would have failed. Then a new primitive concept would have come forward, perhaps one of internal propulsion: that planets have an internal mechanism or guidance system that propels them along their orbits, witness the force emanating from within in the form of volcanoes and earthquakes. Astronomers would sink shafts deep into the earth rather than merely gaze at the heavens.

The primitive concept of intelligence gives rise to a heuristic for guiding theory construction and has no ontological status beyond that. It gives advice to social scientists, that is, it warns us that we omit an 'intelligence factor' from our theories of human behaviour at our peril. There is nothing to be gained from trying to sharpen or measure the primitive concept of intelligence. The thing to do is follow its advice and formulate a scientific construct of a mental ability or abilities other than memory and learning that has explanatory power. Jensen has pursued the right path in worrying little about the 'nature' of intelligence and attempting to exploit the potentialities of *g*: it is *g* that must earn real-world ontological status by explaining the life-histories or achievements of individuals and groups. A lot of time is being wasted asking people about their concepts of intelligence and examining ordinary language. Or better, this research has a purpose but it is purely one of communication rather than advancing explanation. Knowing how people in general use the words intelligence and its derivatives like 'bright' and 'dull' will indicate how far these differ from our scientific constructs and whether we can use such words in describing our results without being irresponsible. It is of course irresponsible to use them where there is any reason to believe our scientific constructs have failed, such as in explaining between-group differences. Even if the link with ordinary language usually exists, it is then broken.

Lack of interest in the primitive concept of intelligence does not entail the dilemmas of crude operationalism. Intelligence is not what IQ tests measure; and therefore, there is no problem about how we can say that present IQ tests are not perfect measures of intelligence. Intelligence is a description of an explanatory role IQ tests aspire to play and therefore, IQ tests must be revised when they do not play that role well — when IQ differences occur and predicted real-world achievement differences do not. When you formulate a theory to explain the life-histories of individuals and groups, your theory will lose explanatory power unless it includes a mental ability or abilities distinct from memory and learning. That cannot be entirely bad advice despite the selective failures of g and the Spearman—Jensen theory.

The failures of the theory are not Jensen's failures. As a working scientist, he has rarely put a foot wrong and he is certainly not guilty of the philosophical mistakes of which he has been accused. IQ tests simply have not given him a good enough g and psychology has not advanced far enough to integrate scientific constructs of intelligence into a more complex and comprehensive theory of personal and social development.

University of Otago
Dunedin, New Zealand

BIBLIOGRAPHY

Bane, M. J., and Jencks, C.: 1977, 'Five Myths About Your IQ', in N. J. Block and G. Dworkin (eds.), *The IQ Controversy: Critical Readings*, Quartet Books, London, pp. 325—338.

Block, N. J., and Dworkin, G.: 1977, 'IQ, Heritability, and Inequality', in N. J. Block and G Dworkin (eds.), *The IQ Controversy: Critical Readings*, Quartet Books, London, pp. 410—540.

Bouvier, U.: 1969, *Evolution des Cotes à Quelques Tests* (Evolution of Scores from Several Tests), Belgian Armed Forces, Center for Research into Human Traits, Brussels.

Brody, N. and Brody, E. P.: 1980, 'Differential Construct Validity', *Behavioral and Brain Sciences* **3**, 335—336.

Cattell, R. B.: 1963, 'Theory of Fluid and Crystalized Intelligence: A Critical Experiment', *Journal of Educational Psychology* **54**, 1—22.

Crano, W. D.: 1974, 'Cognitive Analyses of the Effects of Socioeconomic Status and Initial Intellectual Endowment on Patterns of Cognitive Development and Academic Achievement', in D. R. Green (ed.), *The Aptitude-Achievement Distinction*, McGraw-Hill, New York, pp. 223—253.

Crano, W. D., Kenny, D. A., and Campbell, D. T.: 1972, 'Does Intelligence Cause Achievement? A Cross-Legged Panel Analysis', *Journal of Educational Psychology* **63**, 258—275.

Economos, J.: 1980, 'Bias Cuts Deeper Than Scores', *Behavioral and Brain Sciences* **3**, 342—343.

Elley, W. B.: 1969, 'Changes in Mental Ability in New Zealand Schoolchildren', *New Zealand Journal of Educational Studies* **4**, 140—155.

Eysenck, H. J.: 1985, 'The Nature of Cognitive Differences Between Blacks and Whites', *Behavioral and Brain Sciences* **8**, 229.

Eysenck, H. J., and Barnett, P.: 1985, 'Psychophysiology and Measurement of Intelligence', in C. R. Reynolds and V. Wilson (eds.), *Methodological and Statistical Advances in the Study of Individual Differences*, Plenum, New York, pp. 1—49.

Flynn, J. R.: 1980, *Race, IQ, and Jensen*, Routledge, London.

Flynn, J. R.: 1984, 'The Mean IQ of Americans: Massive Gains 1932 to 1978', *Psychological Bulletin* **95**, 29—51.

Flynn, J. R.: 1985, 'Wechsler Intelligence Tests: Do We Really Have a Criterion of Mental Retardation?', *American Journal of Mental Deficiency* **90**, 236—244.

Flynn, J. R.: 1987a, 'Flynn Replies to Nichols', in S. Modgil and C. Modgil (eds.), *Arthur Jensen: Consensus and Controversy*, Falmer Press, Lewes, Sussex, pp. 234—235.

Flynn, J. R.: 1987b, 'Massive IQ Gains in 14 Nations: What IQ Tests Really Measure', *Psychological Bulletin* **101**, pp. 171—191.

Flynn, J. R.: 1987c, 'Race and IQ: Jensen's Case Refuted', in S. Modgil and C. Modgil (eds.), *Arthur Jensen: Consensus and Controversy*, Falmer Press, Lewes, Sussex, pp. 221—232.

Gibson, J. B.: 1970, 'Biological Aspects of a High Socio-Economic Group: I. IQ, Education and Social Mobility', *Journal of Biosocial Science* **2**, 1—16.

Girod, M., and Allaume, G.: 1976, 'L'Évolution du Niveau Intellectuel de la Population Francaise pendant le Dernier Quart de Siècle' (The Evolution of the Intellectual Level of the French Population during the Last Quarter Century), *International Review of Applied Psychology* **25**, 121—123.

Gould, S. J.: 1981, *The Mismeasure of Man*, Norton, New York.

Green, D. R. (ed.): 1974, *The Aptitude-Achievement Distinction*, McGraw-Hill, New York.

Harlow, H. F., and Harlow, M. K.: 1962, 'The Mind of Man', in *Year-Book of Science and Technology*, McGraw-Hill, New York, pp. 31—39.

Jensen, A. R.: 1972, *Genetics and Education*, Methuen, London.

Jensen, A. R.: 1973a, *Educability and Group Differences*, Methuen, London.
Jensen, A. R.: 1973b, *Educational Differences*, Methuen, London.
Jensen, A. R.: 1979, 'The Nature of Intelligence and Its Relation to Learning', *Journal of Research and Development in Education* **12**, 79—95.
Jensen, A. R.: 1980a, *Bias in Mental Testing*, Methuen, London.
Jensen, A. R.: 1980b, 'Author's Response: Bias in Mental Testing', *Behavioral and Brain Sciences* **3**, 359—368.
Jensen, A. R.: 1981, *Straight Talk About Mental Tests*, The Free Press, New York.
Jensen, A. R.: 1982, 'The Debunking of Scientific Fossils and Straw Persons', *Contemporary Education Review* **1**, 121—135.
Jensen, A. R.: 1985, 'The Nature of the Black-White Difference on Various Psychometric Tests: Spearman's Hypothesis', *Behavioral and Brain Sciences* **8**, 193—219.
Jensen, A. R.: Undated (but circa 1985), 'The *g* Beyond Factor Analysis', unpublished manuscript, courtesy of A. R. Jensen, University of California, Berkeley, California.
Jensen, A. R.: 1987, 'Differential Psychology: Towards Consensus', in S. Modgil and C. Modgil (eds.), *Arthur Jensen: Consensus and Controversy*, Falmer Press, Lewes, Sussex, pp. 353—399.
Keynes, J. M.: 1936, *The General Theory of Employment, Interest and Money*, Macmillan, London.
Leeuw, J. de, & Meester, A. C.: 1984, 'Over het Intelligente-Onderzoek bij de Militaire Keuringen vanaf 1925 tot Heden' (Intelligence — as Tested at Selections for the Military Service from 1925 to the Present), *Mens en Maatschappij* **59**, 5—26.
Lewontin, R. C.: 1977, 'Race and Intelligence', in N. J. Block and G. Dworkin (eds.), *The IQ Controversy: Critical Readings*, Quartet Books, London, pp. 78—92.
Lynn, R.: 1982, 'IQ in Japan and the United States Shows a Growing Disparity', *Nature* **297**, 222—223.
Lynn, R., & Hampson, S.: 1986, 'The Rise of National Intelligence: Evidence from Britain, Japan and the U.S.A.', *Personality and Individual Differences* **7**, 23—32.
Lynn, R., Hampson, S. L., and Mullineux, J. C.: undated (but circa 1986), 'The Rise of Fluid Intelligence in Great Britain, 1935—1985', unpublished manuscript, courtesy of R. Lynn, University of Ulster, Coleraine, Londonderry, Northern Ireland.
McClelland, D. C.: 1977, 'Testing for Competence Rather Than for "Intelligence"', in N. J. Block and G. Dworkin (eds.), *The IQ Controversy: Critical Readings*, Quartet Books, London, pp. 45—73.
Mehlhorn, G., and Mehlhorn, H.-G.: 1981, 'Intelligenz-Tests und Leistung' (Intelligence-Tests and Achievement), *Wissenschaft und Fortschritt* **31** (9), 346—351.
Nichols, R. C.: 1987, 'Racial Differences in Intelligence', in S. Modgil and C. Modgil (eds.), *Arthur Jensen: Consensus and Controversy*, Falmer Press, Lewes, Sussex, pp. 213—220.
Rist, T.: 1982, *Det Intellektuelle Prestasjonsnivaet I Befolkningen Sett I Lys av den Samfunns-Messige Utviklinga* (The Level of the Intellectual Performance of the Population Seen in the Light of Developments in the Community), Norwegian Armed Forces Psychology Service, Oslo.
Scafer, E. W. P.: 1985, 'Neural Adaptibility: A Biological Determinant of *g* Factor Intelligence', *Behavioral and Brain Sciences* **8**, 240—241.

Schallberger, U.: 1985, *HAWIK und HAWIK-R: Ein Empirischer Vergleich* (HAWIK and HAWIK-R: An Empirical Comparison) (Tech. Rep.), Psychologisches Institut der Universität, Zurich.

Spearman, C.: 1904, '"General Intelligence": Objectively Determined and Measured', *American Journal of Psychology* **15**, 201—292.

Sternberg, R. J.: 1985, *Beyond IQ: A Triarchic Theory of Human Intelligence*, Cambridge University Press, Cambridge, England.

Terman, L. M., and Oden, M.: 1959, *The Gifted Group at Mid-Life*, Stanford University Press, Stanford, California.

Thorndike, R. L.: 1973, *Stanford-Binet Intelligence Scale 1972 Norms Tables*, Houghton Mifflin, Boston.

Thurstone, L. L.: 1940, 'Current Issues in Factor Analysis', *Psychological Bulletin* **37**, 189—236.

Tuddenham, R. D.: 1948, 'Soldier Intelligence in World Wars I and II', *American Psychologist* **3**, 54—56.

Vogel, F.: 1980, 'Genetic Influences on IQ', *Behavioral and Brain Sciences* **3**, 358.

Waller, J. H.: 1971, 'Achievement and Social Mobility: Relationships Among IQ Score, Education, and Occupation in Two Generations', *Social Biology* **18**, 252—259.