

The Current Status of the IQ Controversy*

Arthur R. Jensen, *University of California*

The so-called IQ controversy has endured longer than any other controversy in the history of psychology. Normally, scientific controversies wax and wane, and often die out completely. The IQ controversy, however, remains very much alive and, in fact, has intensified in recent years, with no signs of any letup in the foreseeable future.

Scientific controversy can involve dispute over *facts* (i.e., observations, measurements, events, statistical analyses of data) or dispute over *theory* (and the hypotheses that flow from it), or both. Any given controversy must be viewed from the standpoints both of fact and theory.

The normal, or at least desirable, course of events is for scientists to arrive at a consensus as to some of the facts that must be taken account of in a given domain of scientific interest. They then formulate a theory that can comprehend the already established facts and logically and rigorously generate hypotheses that are, in principle, empirically falsifiable; in this way, appropriate tests of a hypothesis can result in its rejection and thereby in the discovery of new facts. The discovery of objective knowledge is the real aim of scientific investigation, not the creation of theories *per se*. (By "objective" I simply mean sensory observations, either unaided or aided by scientific instruments, on which many people making the observations are in general agreement as to what they have observed.) Theories are highly important tools—scaffolding—for acquiring objective knowledge of nature. Theories are among the most important tools of discovery; they lead us to look where we might not have looked otherwise; they highlight relationships that might otherwise go unnoticed; and they sometimes generate far-out predictions that are counterintuitive and violate all common sense. But the main purpose of a theory is to lead us to objective facts we did not know before. Some facts also take the form of being able to *do* certain things that we could not do before. One indicator of the success of scientific endeavor is the undisputed

* This article is based on lectures given by the author at the Universities of Adelaide, La Trobe, Melbourne, and Sydney, in Australia during September-October 1977. The lectures were cancelled by the authorities in three other Australian universities because of threatened demonstrations against the author's appearance on their campuses.

results of its technological application, its power to cause events of practical consequence in the real world, whatever the value judgments that we may make about them.

Thus, controversy over theories, in the sense of deducing hypotheses and testing them by collecting appropriate evidence, and the critical analysis of data and methodology used in testing hypotheses, is *normal science*. It is the healthy state of affairs, and when there is any diminution of controversy in this normal sense, we should begin to worry, for it is a sure sign of a moribund field. In this sense, the scientific aspects of the IQ controversy are today in even better health than at any time in the past history of psychology.

But the IQ controversy, for better or worse, has also come to mean a kind of controversy beyond the realm of normal science. Much of the controversy is not intrinsic to the scientific issues, but is ideological, stemming from differing philosophic ideals, opinions, or sentiments as to which form of social, political, and economic order is most desirable. In this respect, the present IQ controversy has much in common with other great controversies in the history of science, controversies inflamed by philosophic or religious issues not intrinsic to the scientific questions involved, such as surrounded Galileo's claim of Jupiter's moons, Copernicus's heliocentric theory, and Darwin's theory of evolution. In each case, the extrinsic debate falls away in due time and the intrinsic scientific controversies proceed as normal science, making for advances in our knowledge and its practical uses. We can expect the same to happen in the IQ controversy, as long as there are some researchers whose efforts are mainly concerned with the intrinsic scientific aspects, and they are not too harassed, discouraged, or sidetracked by the accompanying ideological polemics. Scientific progress is never made by putting down ideological arguments (which by their very nature are immune to any possible empirical falsification, and hence are outside the pale of science), but only by relentlessly pursuing normal science. Therefore, I think it much more important and interesting to present a synoptic overview of the normal science aspects of the current IQ controversy than a description of the ideological polemics with all their stubbornly unvarying clichés and mantras. That is not to say that the latter could not also be interesting grist for analysis by students of the history and sociology of science, but that is neither my field of competence nor the best use I could make of the limited space available here.

The IQ controversy is actually a number of controversies which, although interrelated, must be examined separately. They can be grouped under four main headings: (1) the nature and measurement of intelligence; (2) the heritability of IQ *within* culturally homogeneous populations; (3) the genetic component in the mean IQ differences *between* groups, that is, between social classes within racially and culturally homogeneous populations, and between

different racial populations; and (4) the social and educational implications of a genetic component in IQ variance. These last issues obviously are not scientific ones, but are matters of educational and social policy, which we should hope will take account of the scientific facts, where these are well established, and remain openly agnostic, where they are not. Policies must be decided in terms of human values, also taking account of available facts, including the cost/effectiveness ratio for different courses of action and their probable outcomes.

I could not possibly go into all of the scientific and policy controversies involved in the whole IQ controversy, or even one of them in any detail, in this short article. I can only hope to indicate in sketchy outline a few of the main points of general agreement and of serious doubt or disagreement among present-day scientists working in this field.

THE NATURE AND MEASUREMENT OF INTELLIGENCE

A wide range of individual differences in ability to perform mental tasks of many kinds is obvious and undisputed. By "mental" I mean that little if any of the population variance in performance on the tasks is attributable to individual differences in sensory and motor functions *per se*. An almost infinite variety of mental tasks can be made up in which there is variance in random or representative samples of the total population of any given age group, and performance on these tasks or items can be objectively scored in terms of right or wrong, pass or fail, or the time taken for successful completion. Also, such tasks can be ordered in difficulty, when difficulty is defined as the proportion of the population attempting the item whose performance on that item is scored "pass". These, then, are the undisputed "givens" of the situation.

The first important observation for any theory of mental ability is the well established fact that practically all such mental test items are positively intercorrelated in representative samples of the general population. The positive intercorrelations extend over a phenomenally diverse collection of items, such that there is apparently no limit, except that set by our own ingenuity, to the number of different items that will correlate positively with all the others in this domain.

This matrix of positive intercorrelations among mental test items logically implies that all of the items that show significant correlations with all of the other items must be measuring at least one source of variance in common. This is what is meant by the construct of a general factor common to all mental tests. Charles Spearman labeled it *g*. The existence of *g*, at the level of correlational analysis, is now an established fact. The invention of factor analysis permitted the estimation of the amount (in the sense of proportion of variance accounted for) of the *g* factor in any given test or test item when

factor analyzed among a collection of other items. The *g* factor (which may also be described as the first principal component of the intercorrelation matrix) cannot be eliminated by any mathematical technique of factor analysis, no matter how hard one may wish to try. Orthogonal rotation of the factor axes may at most obscure the general factor; the fact that Thurstone's criterion of "simple structure" cannot be approximated by orthogonal rotation of factors in the abilities domain is only further proof of the existence of *g*.

In any large and highly diverse battery of mental tests or items, we can usually identify several other factors (so-called "group factors"), such as verbal, numerical, spatial, and memory, which all together usually account for less of the total variance than the *g* factor alone. Other smaller group factors are usually ephemeral from one factor analysis to another, and of little practical importance.

The large *g* factor of an extensive and diverse battery of mental test items thus can be justifiably viewed as an operational definition of intelligence. It highly accords with common sense notions of intelligence, but is also more precise and uncontaminated by sources of variance (e.g., motivation, personality, etc.) which are not correlated with *g*, but which often distort one's subjective judgments of persons' intelligence. Few of us have any trouble recognizing differences in capability between persons at the two extremes of mental ability; individuals at the extremes of what most people commonly think of as intelligence also differ most markedly on psychometric measures of *g*, more than on measures of anything else. Between these two extremes of ability there is a smoothly graded continuum of individual differences in the trait we label as *g*. There is good reason to believe that *g* has an approximately normal or Gaussian distribution in the interval of about ± 2.5 standard deviations from the mean of the general population.

The kinds of test items most heavily loaded with *g*, as revealed by numerous factor analyses in the past 70 years, are those involving some mental complexity: reasoning, problem solving, and conceptual and semantic discrimination and generalization. But the *g* factor is involved whenever any kind of mental manipulation or transformation must be performed on the stimulus input in order to arrive at the required solution or response, and the more complex the manipulation or transformation, the more *g* loaded is the item. I have expanded on this topic elsewhere (Jensen, 1978b).

An intelligence test, then, is (or should be) a collection of quite highly *g* loaded items with enough diversity to balance out the smaller group factors, so that most by far of the variance in total scores on the test is attributable to *g*. The total scores on most standardized tests of general intelligence in current use show *g* loadings (i.e. correlations with *g*) of about .75 to .90 when factor analyzed among a large collection of diverse mental tests. Such a collection of *g* loaded

test items, when standardized by age groups in representative samples of some clearly specified population, is often called an IQ test, where the IQ is a standardized score, traditionally with a mean of 100 and standard deviation of 15 at every age level.

Whether or not any particular mental test item can legitimately be said to measure intelligence for members of a given population depends entirely on a wholly objective criterion, viz., the size of the item's g loading when it is factor analyzed among many other items in a representative sample of persons from that population. If it is not significantly loaded on g (i.e., the first principal component), it does not measure what we mean operationally by intelligence and should be excluded from any test so labeled.

Highly g loaded tests show substantial correlations with a host of educationally and occupationally important criteria, and have often been of demonstrated practical value in scholastic and vocational prediction, guidance, and selection.

Notice that g has not been "reified," here, as some critics of the construct are prone to complain. The g factor is acknowledged as a purely mathematical, theoretical construct needed to account for the raw fact of overwhelmingly positive intercorrelations among mental tests.

At present there is no well developed theory of the underlying nature of g , other than to descriptively characterize the kinds of test items that are the most and the least g loaded. But that was done quite well by Spearman more than fifty years ago. The major advances we most need to make now involve investigation of the anatomical, histological, physiological, biochemical and electrochemical underpinnings of g . Our knowledge in this realm is still rudimentary, but future theories of intelligence will have to account for such already well established facts, for example, as the correlation (of about +.30) between brain size and IQ (Van Valen, 1974), and between the amplitude and latency of EEG visual and auditory evoked potentials and IQ (Callaway, 1975).

Development of theories of this type will most likely have to appeal to biological thinking, linking human intelligence to evolutionary theory and the comparative psychology of abilities in lower animals (Viaud, 1960).

In short, there seems to be little, if any, serious controversy at present about the broad facts of g as a theoretical construct: we can reliably measure g in individuals, and g has significant correlations with many educationally, vocationally, and socially important variables. Counterarguments—without any real empirical or theoretical substance—are occasionally directed against this generally accepted position (almost always in the popular media and rarely in the technical literature), usually as a first line of attack on the idea that individual or group differences in intelligence are hereditary. For

example, it has been claimed by some critics, without supporting argument, that since the IQ is not an absolute scale (i.e., a scale with an absolute zero) the methods of quantitative genetic analysis cannot legitimately be applied to IQ or other mental test scores which are only interval or ordinal scales (Jensen, 1975d, p. 177). Since quantitative genetics is based entirely on correlation and regression analysis and analysis of variance, in which the grand mean of all the measurements is irrelevant, an absolute scale of measurement is quite unnecessary, however desirable it might be for other purposes, such as for the study of the growth of intelligence from infancy to maturity.

The current scientific controversies involving the inheritance of intelligence take the construct of *g* and its usefully reliable measurement by means of IQ tests as sufficiently well established to provide an empirical basis for investigation of the inheritance of intelligence by the methods of biometrical genetics.

THE HERITABILITY OF INDIVIDUAL DIFFERENCES IN IQ

One of the main classes of phenomena that any comprehensive theory of intelligence must explain is the highly distinctive pattern of correlations found between the IQs of persons of varying degrees of genetic kinship, such as monozygotic and dizygotic twins, siblings, parent-child, cousins, and genetically unrelated children reared together (Erlenmeyer—Kimling & Jarvik, 1963). There are now numerous studies of IQ correlations for various kinship groups, and the measures of central tendency of the various kinship correlations have converged on quite stable values which are distinctly different for different degrees of genetic kinship and also according to whether persons of a particular kinship were reared together or apart. The median correlations between IQs are highest for the closest degree of kinship (monozygotic twins), and decrease systematically in a clear stepwise fashion for lesser degrees of kinship, such as monozygotic and dizygotic twins, are considerably greater than the correlation differences between persons of the same kinship who have been reared together or reared apart. Differences in kinship produce larger differences in correlation than differences in the conditions of rearing (i.e. reared together or reared apart). Also, the IQs of adopted children are much more highly correlated with the intelligence levels of their biological parents, with whom they have had no contact since early infancy, than with the intelligence levels of the adoptive parents who have reared them from infancy (Munsinger, 1975).

One can quibble about certain methodological shortcomings in this or that particular kinship study, but there are now many studies of all the major kinships and they are most valuable for biometrical genetic analysis. It is difficult to argue about the median values of all these kinship correlations reported in the literature—which, in-

cidentally, has been rapidly growing in recent years. There is here clearly an imposing and distinctive class of phenomena that needs to be explained.

To date, no strictly psychological theory has been formulated that can begin to comprehend this set of facts. However, they are generally predicted by a polygenic model originally developed in the field of quantitative or biometrical genetics by such pioneers as R. A. Fisher, Sewall Wright, J. B. S. Haldane, and Kenneth Mather. The polygenic model of inheritance is applicable to all metrical characteristics in animals and plants. I cannot here go into the technical aspects of the polygenic model or the statistical methodology involved in its empirical testability. It is not a finished and static affair, but is presently in a state of scientific ferment, new developments are taking place, as well as new technical criticism and analysis (for an introduction, see Jinks & Fulker, 1970.)

Nearly all of this activity has originated within the fields of biometrical genetics and behavioural genetics, in which the most prominent leaders in recent theoretical and methodological developments are geneticists John F. Jinks and Lindon Eaves, in the University of Birmingham, England, and Newton Morton and colleagues in the University of Hawaii. (See Eaves, 1975; Eaves, Last, Martin & Jinks, 1977; Morton, 1974.)

The current ferment in this field is a sign of scientific vigour. Yet outsiders might easily get the false impression that the "genetic theory of IQ" is under fatal siege. One hears that this or that prominent geneticist criticized or questioned this or that point in some aspect of the vast topic, as if to imply that the whole of our knowledge in this field is about to tumble at last! Far from it. Among the experts in this field there is, in fact, very general agreement that IQ variation involves genetic factors and that any comprehensive and detailed scientific understanding of the phenomena of IQ variation, such as I have outlined above, will have to involve genetic principles and formulations (Jensen, 1975a).

The *normal science* controversies in this field now are not concerned with whether IQ variation has a substantial genetic component; instead, they have to do with the development of refinements and elaborations of the polygenic model to take account of several different, more complex, sources of genetic and environmental variation, such as dominance deviation and epistasis, assortative mating, and variance due to covariance and the interaction of genotype and environment (e.g., De Fries, Vandenberg & McClearn, 1976; Plomin, De Fries, & Loehlin, 1977; Jensen, 1978a). These extensions and refinements of the basic polygenic model—if they are to be scientifically fruitful—necessitate the development of more powerful methodologies for empirically testing the model, and this constitutes a prominent part of the recent technical activity in the field. These developments in

biometrical genetics are applicable in principle not only to the study of IQ and other abilities, but to all behavioural traits. Here, as is true of the growing frontier of every science, there are many technical problems and disagreements, of such a highly specialized nature as to be scarcely understandable to persons who are not well versed in biometrical genetics and multivariate statistics. The unresolved problems, which are inevitable in any developing sphere of science, are a challenge to the specialists in quantitative genetics who are interested in extending their models to account for complex human behaviour.

But the question of IQ heritability in the broad sense, is no longer a serious topic of controversy among workers in this field. Most would agree that the lower probable limit of the broad heritability of IQ is about .50 (i.e. $>50\%$ of the IQ variance is attributable to variance in genotypes), and most estimates fall in the range from about .65 to .80. (Heritability is not a universal constant, but a population statistic, and so no one is concerned with trying to determine the "true" or "exact" heritability of IQ in general; that is an inappropriate aim). A broad heritability of .65 to .80 implies a correlation between individuals' measured IQs and their genotype for the development of intelligence of about .80 to .90 (i.e. the square root of the broad heritability). It is the further detailed analysis of that broad heritability which is now of central interest, as well as the developmental genetic aspects of the mechanisms by which nongenetic or environmental factors influence the phenotypic expression of the genotype.

The Burt Controversy. Sir Cyril Burt, who died in his 89th year in 1971, was a pioneer in the study of the genetics of mental ability, having been the first psychologist to introduce the polygenic models of Fisher and Mather in this field, along with masses of relevant kinship data on IQ which Burt had collected over the years in the London schools. After Burt's death, I pulled together all of the published results of his studies on the genetics of mental ability and systematically arranged all of the various kinship correlations in a series of nine large tables (Jensen, 1974a). When the whole of Burt's reported data and kinship correlations were thus arrayed, certain peculiarities (certainly errors of some kind) became apparent in the figures. About a year before my 1974 article appeared, Leon Kamin, in a psychological convention address (also in Kamin, 1974), had brought attention to at least one of the many anomalies in Burt's reporting of his twin studies: the repetition of the exact correlation of .771 (for identical twins reared apart) in three different articles spaced several years apart, with reported sample sizes of 21, 30 and 53 twin pairs. (Burt was not explicit about the N of 30, however, so we are sure of only one repetition of the same correlation with different N s.) In my 1974 article, I turned up a total of no less than twenty similar anomalies, where the N changes from one report to the next, but the corresponding kinship correlation remains the same

to three decimal places. But all twenty of the "constant correlations" are linked to only eight of the kinship samples (out of a total of 48) reported by Burt, since usually kinship correlations for several different mental tests and physical measures were determined on the same sample. Burt went on cumulating kinship data throughout his long career, and after 1955 the numerical anomalies in his reports of these studies seem to compound. Several psychologists have charged Burt with out-and-out fraud, faking data, or sheer invention of results to fit his polygenic model of intelligence, and it was even claimed that two of Burt's research collaborators and co-authors were fictitious (Gillie, 1976).

I have written elsewhere in some detail about the flimsiness of these charges against Burt (Jensen, 1977a; 1978c). Now, more than two years since these defamatory charges were made, there is still no substantiating evidence that has come forth.

Several of the anomalies in Burt's figures are rather transparent copying errors, such as reversing, transposing, or substituting digits. Overall, these errors are unsystematic and seem most likely due to carelessness in copying and proofreading tables. This degree of carelessness and frequent omissions of important details, such as *Ns* and *SDs*, in the reporting of research results is itself a serious offence for a scientist, and stands in puzzling contrast to Burt's elegant style of writing, his high level of technical sophistication in genetics and statistics, and the extreme rarity of theoretical or conceptual errors in his work. It is interesting to note that the rate of numerical errors in the journal references found in the bibliographies of Burt's articles is about the same as the rate of numerical errors in his reported sample sizes and correlations (McAskie, 1978). So while there is *prima facie* evidence for careless numerical errors in reporting results, there as yet has been no evidence for fraud. A chief suspicion that remains unresolved, however, arises from the apparent difficulty in accounting for the addition of 32 pairs of identical twins reared apart to Burt's twin collection between the years 1955 and 1966, when Burt was between 73 and 83 years of age. (Also, in private correspondence he claimed to have added to his collection, after 1966, three more pairs of MZ twins reared apart, bringing the total to 56 pairs [Sandra Scarr, personal communication].) The answer, if it can ever be found, will now probably have to depend on indirect inferences from careful biographical research; such investigation is being conducted by Professor Leslie Hearnshaw (University of Liverpool), a noted historian of British psychology, who is presently at work on a full scale biography of Burt.

Whatever further investigation may eventually reveal about Burt or his data is now only of historic and biographic interest. It has virtually no scientific relevance today, since the kinship correlations of greatest value in genetical research have been amply replicated by

many other investigators both before and after Burt's publications. The total deletion of Burt's now questioned empirical legacy would at this time scarcely make an iota of difference to any general conclusions regarding the heritability of intelligence, so much greater is the body of more recent and better evidence. Thus, from a scientific standpoint, the case of Burt is only an interesting biographical sidelight, rather than an intrinsic element of the IQ controversy itself.

A GENETIC COMPONENT IN MEAN DIFFERENCES BETWEEN GROUPS

There are two main issues here which have quite different statuses, scientifically, and must be dealt with separately. The first is the question of social class differences in IQ. The second is the question of racial differences.

Socioeconomic Status (SES) and IQ

The positive correlation between indices of SES and IQ (as well as all of the correlates of IQ) is well established (Jensen, 1972a, 1973a). The correlation is considerably higher for adults, who have already attained their educational, occupational, and socioeconomic level, than for children, whose SES must be based on the status of their parents. There is considerable social mobility between generations in most modern industrial societies, and some sizeable percentage of siblings move above, and some fall below, the SES of their parents. Such social mobility is closely linked to success in school, amount of education, and eventual occupational level. It should not be surprising, therefore, that SES is positively correlated with IQ. The process by which such a correlation should be expected to come about is highly apparent.

If individual differences in IQ involve genetic factors as well as environmental, as the research on IQ heritability indicates, then it is extremely implausible that individual differences in scholastic performance and occupational attainment would be correlated only with that component of IQ variance which is contributed solely by environmental effects. The clincher is that the correlation between IQ and SES also exists *within* families, that is, among siblings who were reared in the same home and who therefore do not differ in social or cultural background. For example, sons whose IQs are higher than their fathers' IQs (when both are tested at the same age in school) attain by middle-age a higher SES level, on the average, than their fathers had attained at the same age; and sons with lower IQs than their fathers' attain a lower SES (Waller, 1971). Also, the IQs of orphanage children are correlated with the occupational levels of their fathers, of whom they have had no knowledge (Lawrence, 1931). Such findings make it virtually impossible to conclude other than that the observed correlation between IQ and SES, in school-age children and

adults, involves a substantial genetic component. The compelling evidence for this conclusion has practically made SES differences in IQ a noncontroversial subject among nearly all geneticists, psychologists, and sociologists who have studied the matter.

Race Differences in IQ

This is the real crux of the IQ controversy. It generates nearly all of the heat and fumes, which have been most pronounced in the popular media. Much of the resistance to research on the heritability of individual differences in mental abilities is, I believe, motivated by fear of any implications the findings might have for the explanation of racial differences in IQ and all of its socially significant correlates, in which there are also conspicuous racial differences. The fact is that IQ is correlated significantly with highly valued educational, occupational, and socioeconomic criteria to about the same degree both *within* and *between* various racial groups (at least in the United States). That is to say, in the U.S.A. whites and blacks of comparable IQ show comparable educational and occupational attainments, and comparable rates of delinquency and criminal offenses. Were it not for these socially important correlates of IQ both within and across racial groups, I believe the concept of IQ would be much less attacked than it is. There would be little fulmination over the measurement of any characteristics showing individual differences and racial differences that have no obvious socially important correlates. Scientists have studied race differences in blood pressure, without being harassed or denounced as "racist". Yet the scientific problems of studying the genetics of racial differences in blood pressure are remarkably parallel to the IQ question. Blood pressure is a metric characteristic which shows substantial heritability, but is also affected by dietary habits and environmental stresses, in which there are both individual and group differences.

The race-IQ controversy involves a number of themes on which current methodology, research evidence, and consensus of informed opinion have advanced to quite different degrees.

1. Test bias. Perhaps the most fundamental question that needs to be answered is whether the observed racial differences in IQ can be explained in terms of some form of bias in the IQ tests themselves. The crucial question is not whether *some* tests are culture biased. For even if no existing tests were culture biased, it should be entirely possible to construct a culture biased test, if this concept has any real meaning. Nor is it a crucial question whether a given existing test is culture biased with respect to the comparison of two particular cultural groups. Let us grant, as axiomatic, that for any given test (or any possible test), at least two cultural groups can be found for which the test will be culture biased.

The crucial question is this: Is there any mental test that measures *g* and that shows a difference between two racial groups which cannot be attributed to test bias *per se*? It should be obvious that a significant mean difference in test scores between the racial groups cannot itself constitute evidence of test bias, for that is the very point in question. Nor are subjective judgments of whether a test (or a single test item) is biased a scientifically admissible criterion. Experiments have shown that psychologists' intuitive judgments of which particular tests or test items will discriminate the most or the least between two racial groups are extremely fallible and show little agreement among the judges (Jensen, 1977b). We need to use objective psychometric criteria of bias that are amenable to statistical hypothesis testing. The critics of mental testing who claim that IQ tests are culturally biased and go on to explain racial differences in IQ on this basis never provide any *independent* objective evidence that the IQ tests are biased. In fact, it seems safe to say that no objective criteria of bias have ever been produced by those who invoke culture bias as an explanation of observed racial differences in test scores.

Yet there are a large number of objective criteria by which we can recognize test bias. First, it should be established that the main source of variance in total scores (i.e., the *g* factor or first principal component) is the same factor in both racial groups. This can be done by comparing the relative magnitudes of factor loadings on the various tests or items across the groups. Then we can go on to compare the two groups on such psychometric indices as the following: internal consistency reliability and item-total correlation, predictive validity for scholastic or occupational performance or other criteria, correlation of raw scores with chronological age, item characteristic curves, the items \times race interaction in the analysis of variance of the item-score matrix, the rank order of item difficulties, the relative frequencies of errors on the various multiple-choice distractors, and the average absolute difference among full siblings reared in the same family. Each of these features of the test data provides an objectively testable criterion of bias, in terms of whether these psychometric properties of the test behave the same or differently in the two groups under comparison. These criteria, of course, are not above criticism. Their chief virtue is that they are objective and therefore can be debated theoretically and tested empirically like any other statistically framed hypotheses. Anyone can add his own objective criterion of bias to this list. Probably no single criterion is crucial; but in combination they are like a series of sieves for detecting bias.

When these criteria of bias have been applied to a variety of widely used standardized individual and group intelligence tests, both verbal and nonverbal, they do not reveal any appreciable or directionally consistent bias with respect to American-born blacks and whites (Jensen, 1974b, 1976a, 1977b, 1977c). The same conclusion is

probably true as well for other native-born, English-speaking racial or ethnic minorities in the U.S.A., although the evidence for these groups is not yet nearly so massive or statistically compelling as the evidence on black-white comparisons.

It is also quite clear by now that the black-white IQ difference cannot be attributed to examiner bias. Numerous studies have shown that the race of subjects \times race of examiner interaction is negligible (Jensen, 1974c).

Nor can the differences be attributed in any part to differences in sensorimotor abilities, speed of work, effort, or willingness to comply in a test situation. Tests of each of these factors reveal no appreciable or consistent racial differences (Jensen 1973a).

Therefore, I doubt that the question of test bias *per se* will continue to figure prominently in the IQ controversy as regards racial differences. It is already quite apparent to most researchers in this field that the concept of culture bias in tests will play little part in any scientific explanation of the IQ difference between blacks and whites in the U.S.A. In fact, there has been no objective empirical demonstration of *cultural* differences between the majority of American-born whites and blacks, and what evidence we have even contradicts the notion that whites and blacks in the U.S.A. are culturally different. They differ in average socioeconomic status, but that is not a cultural difference.

The concept of culture bias, of course, is still highly relevant to the use of tests in true cross-cultural studies, where the groups differ in language, customs, values, and almost their entire way of life. I doubt, therefore, that cross-cultural studies are a fruitful approach to reducing the heredity-environment uncertainty regarding racial differences in intelligence. Yet even authentic cross-cultural testing has resulted decisively in the rejection of one popular hypothesis about test bias, namely, the notion that IQ tests inevitably favor the population in which they are standardized or of which the test constructors are members. This is absolutely contradicted by the findings that Arctic Eskimos and Japanese (in Japan), for example, score as high or higher on certain intelligence tests than white Americans, even though the tests were designed by white Americans and Englishmen and standardized for those populations.

2. *Nonexistence of races.* A wholly spurious escape hatch used by some in the IQ controversy is the claim that since there are no races there can be no legitimate question of race differences. The Platonic notion of a pure race is a theoretical fiction with no real importance to the issue (Jensen, 1973b, pp. 342–381; Eysenck, 1971). Different races are viewed biologically as any populations that differ in the frequencies of one or more genes (see Loehlin *et al.*, 1975, Ch. 2). Thus, genetically speaking, race is a quantitative, not a qualitative, variable. The first

essential requirement of a study of racial differences is that the groups involved be random or representative samples of two clearly specified populations which can be shown statistically to differ in the frequencies of one or more genes. The major racial subdivisions, of course, differ in a great many genes.

Another related spurious debating point in the IQ argument is based on the claim (which is true) that human races differ in only some small fraction of any random sample of all their genes. But the next step in the argument is a nonsequitur: that races therefore cannot really be very different. Race differences, of course, do not appear very large when viewed against a background of species differences. (Yet humans even share a great many genes in common with the lower primates.) Despite the overall much greater degree of genetic commonality than of genetic difference among human races, there are obvious genetic differences in many characteristics, which are most apparent at the phenotypic level of polygenic organized systems or traits, as contrasted with the relatively small differences seen in any collection of single genes sampled at random from a large number of loci. Thus, human races show rather easily ascertainable differences in a host of characteristics such as body size and proportions, hair form and distribution, head shape and facial features, cranial capacity and brain formation, blood groups, number of vertebrae, genitalia, bone density, fingerprints, basic metabolic rate, average blood pressure, temperature, heat and cold tolerance, sweating, odor, consistency of ear wax, number of teeth, age of eruption of permanent teeth, fissural patterns on the surfaces of the teeth, length of gestation period, frequency of twins, male-female birth ration, degree of physical maturity at birth, infant development of brain waves, colorblindness, visual and auditory acuity, ability to taste phenylthiocarbamide, intolerance of milk, galvanic skin response, chronic diseases, susceptibility to infectious diseases, and pigmentation of the skin, hair and eyes (Baker, 1974). It seems highly plausible that there would also be racial differences in some behavioural traits that are linked to physical properties of the central nervous system.

One important behavioural difference in which highly consistent differences, both in direction and magnitude, have been found between certain races is in measures of intelligence, that is to say, in highly *g* loaded tests of whatever variety. It should be noted, however, that the largest average differences between representative samples of any races measured on *g* loaded tests, when their schooling has been at least roughly comparable, is only something between one-fifth and two-fifths of the total range of variation encompassed by 99 percent of the members of any large population. But this observation should not obscure the quite considerable consequences of average population differences of one standard deviation or more in *g* type abilities in any educational or socioeconomic system that encourages competition

and selection for *g* loaded activities, such as going to college and entering more desirable occupations (Jensen, 1975b; 1976b).

3. Heritability within and between groups. The broad heritability (i.e., the proportion of total variance attributable to genetic factors) of IQ is about the same *within* the white and black populations of the U.S. (Loehlin, *et al.*—, 1975, Ch. 5). *The big question is: does the fact of the substantial heritability of IQ within racial groups have any implications for the question of a genetic component in the mean difference between racial groups?* I cannot here go into all of the technical aspects of this issue; they are discussed at some length by Jensen (1973a, Ch. 5) and Loehlin *et al.* (1975, Appendix G). Briefly, the answer boils down to this: although the between groups heritability (BGH) can be formally or mathematically related to the within groups heritability (WGH), the formulation is empirically useless, because it contains a parameter (the genetic [intra]class correlation for the specific trait in question among persons within each racial group) which is just as unknown as the BGH. Thus the mathematical relationship between BGH and WGH is represented by a simple equation with two unknowns, and therefore it cannot be solved. That is all that is meant when it is said that there is no necessary or logical connection between WGH and BGH. And that claim is perfectly correct. It is a fact which apparently affords considerable comfort to those who wish to avoid seeking a scientific explanation of the white-black differences in IQ.

On the other hand, it can also be argued, in a way perfectly consistent with probabilistic inference as generally practiced in the physical sciences, that the evidence for WGH probabilistically implies BGH when there is a phenotypic mean difference between the groups, unless it can be demonstrated that the phenotypic mean difference is entirely the result of some other (nongenetic) factor(s). In other words, the inference of BGH from WGH is justifiable unless there is evidence that the causes of differences between groups are essentially different from the causes of differences within groups. The BGH hypothesis cannot be weakened in the least by an *ad hoc* hypothesis of some nongenetic factor for which the only evidence is the mean racial difference in the IQ itself. Yet it is just this kind of *ad hoc* hypothesis that those who insist upon a purely nongenetic explanation of racial differences in IQ have been forced to put forth, since they have been unsuccessful in actually demonstrating any environmental factor (or combination of such factors) that can wholly account for the observed difference. The well known environmental factors that once were commonly believed sufficient to account for all of the white-black IQ difference—factors such as inequalities in schooling, socioeconomic status, nutrition, cultural values, and test bias—have been investigated and found generally inadequate to carry the full burden

of explanation (Jensen, 1973a; Loehlin *et al.*, 1975). Thus, purely environmental explanations are in the position of having to explain quite large race differences in IQ by very weak causal factors, as judged by the effects of these factors on IQ *within* races. It is conceivable that there could be some environmental factor has been substantiated by any evidence independent of the IQ gap itself. A philosopher of science, Peter Urbach (1974), has presented a detailed analysis of the degeneration of the environmentalist position into a hodgepodge of inconsistent *ad hoc* theorizing.

Most researchers in behavioural genetics are now methodologically sophisticated enough to know that certain types of studies, once popular, actually throw remarkably little light on the main issue. For example, consider all the studies that equate racial groups statistically (or by direct matching) on various educational and socioeconomic variables to see by how much the mean IQ differences between the groups will be reduced by controlling these other correlated variables. Such studies cannot reduce the heredity-environment uncertainty unless we make the naive and untenable assumption that the genetic component of IQ (within each racial group) is completely uncorrelated with the statistically controlled or matched variables. This "Sociologist's fallacy," as I have named it (Jensen, 1973a, p. 235), consists of interpreting the results of such studies as if they were true experiments, that is, as if individuals were assigned at random to different environmental conditions, thereby wiping out any significant correlation between genotypes and environmental backgrounds.

We can only approach experimental control to some probably fruitful degree in this field by studying cross-racially adopted children and persons varying in known degrees of racial hybridity but of similar environmental background. The difficulties and ambiguities of non-representative sampling, confounding variables, and other such "escape hatches" for arguments on either side due to methodological deficiencies are exemplified in the few existing studies of these types (see Loehlin *et al.*, 1975, pp. 116–133; McNemar, 1977; Nichols, 1977; Scarr & Weinburg 1976). In these studies, the major findings, where they are theoretically predictable, generally come out in the direction one should expect from a genetic hypothesis, but they are either statistically nonsignificant because of too small samples to provide a strong test of the null hypothesis, or methodological deficiencies and statistically irreparable confounding of variables make the results much like a Rorschach inkblot when it comes to interpretation. Such studies as have already been done, however, are still valuable contributions, not so much for the necessarily weak conclusions that anyone may try to draw from them, but for pointing the way to methodologically stronger studies.

So what we are left with, at present, is merely the considerable *plausibility* of there being some nontrivial genetic component in the

IQ differences between certain racial groups. I am not aware of anyone who claims that environmental factors are not involved to some extent in the average white-black IQ difference, and I have shown evidence for a quite large environmental influence on black IQ in an impoverished population in rural Georgia as contrasted with a relatively affluent community in California [Jensen, 1974d, 1977d].

To be sure, plausibility alone is a scientifically unsatisfactory condition compared with the normal science process of successively testing and revising hypotheses. But as long as strong plausibility remains, we are justified—indeed compelled, if we believe the question is of any importance—to seek ways for attacking the question by the process of normal science. That means entertaining *competing* hypotheses and trying to convert *plausible* hypotheses into *testable* hypotheses. That is all I have ever insisted upon, that we try to go about the job of normal science on this difficult topic. I have deplored the doctrinaire and dogmatic answers and attitudes we so often see on this topic, and which have been so vehemently displayed in recent years by those who, for either sentimental or ideological reasons, apparently abhor the possibility of genetic differences in IQ and its many correlates. They would shun the process of normal science, without which these questions could never be answered objectively. If that be anyone's wish, let him announce it explicitly, so that we can then agree to disagree on fundamental premises and each go our separate ways.

Although plausibility does not constitute scientific knowledge it is a most useful guide to scientific research. We should encourage bringing into focus a much wider range of phenomena with some plausible bearing on the race-IQ controversy than we have been accustomed to looking at in recent years. We especially should take a better look at some of the old, but never adequately researched, questions in this field that in modern times have become almost taboo. As an example, in noting the evidence for a positive correlation between brain size and IQ within the white population (Van Valen, 1974), we might ask if this has any implication for the correlation between mean IQ differences between races and the observed differences in their mean cranial capacities. Juxtaposing certain observations like this can often lead to more detailed elaborations of a plausible theory, thereby giving it more specifically testable facets. Oxford Zoologist John R. Baker (1974) seems to have moved further in this direction of looking for broad relationships between cranial capacity, assessments of indigenous civilizations, and average performance on *g* loaded tests than most behavioural scientists have yet ventured to explore. There can be no tabooed *questions* in science. Reluctance to examine any plausible relationships between possibly relevant phenomena, however "sensitive," can only spread a corrupting influence through science, and I fear this has already happened to

some extent in the behavioural sciences. To preserve silence or maintain an illusion on any one topic creates the necessity for silence or illusion on related topics, and so on, in a widening network of deception and research taboos.

EDUCATIONAL AND SOCIAL IMPLICATIONS

Research findings can have implications for social policies and practical applications only in relation to certain goals and values of the society. These implications do not flow directly from the scientific facts themselves. Scientists have no special qualifications in this sphere, yet they are often called upon for their opinions, and in such circumstances I have tentatively expressed my own opinions concerning the broader educational and social implications of the so-called IQ controversy (Jensen 1972a, pp. 327–332; 1975b; 1976b.) Others have provided further interpretations, which I commend (Bereiter, 1969, 1976; Havender, 1976a, 1976b). Although this is a much too broad and open-ended topic to even begin to do justice to in this paper, I will merely try to summarize my position in two main points.

First, it seems clear that the well established finding of a wide range of individual differences in IQ within all major racial populations and the great amount of overlap of their frequency distributions contradicts the racist philosophy that individuals of different races should be treated differently, one and all, only by reason of their racial differences. Those who would accord any treatment to individuals solely by virtue of their race will find no rational support from any of the scientific findings or theories of modern differential psychology. Man's genetic nature insures individuality, and any doctrine that is built on a denial of this fact is simply at odds with reality.

One's concept of justice derives more from one's moral philosophy than from scientific knowledge *per se*. My concept of justice requires that the fact of statistical differences between racial *populations* should not be permitted to influence the treatment accorded to *individuals* of any race—in education, employment, legal justice, and political and civil rights. This flatly anti-racist philosophy is, of course, a two-edged sword. Righting the past wrongs of racial discrimination can be accomplished best, I believe, by prohibiting racial discrimination in any form, by legal sanctions when necessary, and by seeking equal educational opportunities for members of minority groups who have been denied them in the past, so they can compete fairly in selection for employment, technical training, or higher education, without condescending dispensations.

Second, I believe large individual differences in *g* are here to stay for at least a very long time, and a system of schooling that enforces attendance by virtually the entire society, from age five to sixteen or eighteen years, will have to take fuller recognition of this wide range

of mental ability than has yet been evinced by our educational system. The traditional scholastic curriculum and the usual methods of instruction have proven to be a source of undue frustration and defeat, often leading to overt hostility and school vandalism, for many children of weak academic aptitudes. This is still the great unsolved problem of American public education, a problem that cuts across racial lines and would exist even if there were no racial differences in scholastic performance. The fact that there are racial differences in scholastic performance only makes the problem more difficult to deal with socially and politically. Short of some innovative miracle that could wipe out the high correlation between individual differences in *g* and scholastic achievement, enforced universal public education, if it is to survive at all, is bound to move towards a greater diversity of ways that children can substantially benefit from their years in school. The great difficulty in this is reconciling our ideal of equality of educational opportunity with what is perhaps the necessarily great diversity of educational curricula and goals for different children according to their individual aptitudes. I believe there is a much greater awareness among educators today of the substantial nature of individual differences than in the 1950's and '60's, when it was still hoped that some rather simple environmental manipulations called "compensatory education" could greatly enhance the scholastic performance of many children from groups with traditionally poor performance (Jensen, 1972a). But racial and social class differences do not appear to be essentially different from individual differences in this respect. This is what we should expect, however, if genetic factors are largely responsible for individual and group differences. A substantial heritability means that merely reallocating existing socioeconomic and educational or other existing environmental advantages or disadvantages will not produce any appreciable change in the rank order of individuals or group means, whatever effect such environmental changes might have on the overall population mean. The immediate social problem of education, however, is not generally perceived as the problem of raising the overall absolute level of education attained by the whole population, but as the difficulties of coming to grips with the conspicuous differences in educational attainments that exist among certain groups within the population. As long as that particular issue remains a major social concern, the genetic question is constantly inevitable. Since we are still far from a scientific consensus as to the causes of some of these group differences in educability, the only intellectually warranted official position of educators and governmental policy makers must be one of *open agnosticism* as to the causes, rather than the doctrinaire naive environmentalism that has so long prevailed as official policy (Jensen, 1972b). If scientific agnosticism is deemed unsatisfactory as a permanent state of affairs, and scientists are drawn to the challenge of reducing the heredity-environment uncertainty, they have no choice

but to continue the pursuit of normal science in the IQ controversy. In the history of intellectual conquest, agnosticism concerning socially important natural phenomena has always been a highly unstable condition; it invariably gives way either to dogmatic belief or to scientific knowledge.

Arthur R. Jensen,
Institute of Human Learning,
University of California,
Berkeley.
U.S.A.

REFERENCES

- Baker, J. R. *Race*. New York: Oxford University Press, 1974.
- Bereiter, G. The future of individual differences. *Harvard Educational Review*, 1969, 39, 162-170.
- Bereiter, C. IQ differences and social policy. In Ashline, N. F., Pezzullo, T. R., & Norris, C. E. (Eds.) *Education, Inequality, and National Policy*. Lexington, Mass: Lexington Books, 1976.
- Callaway, E. *Brain Electrical Potentials and Individual Psychological Differences*. New York: Grune & Stratton, 1975.
- De Fries, J. C., Vandenberg, S. G., & McClearn, G. E. Genetics of specific cognitive abilities. *Annual Review of Genetics*, 1976, 10, 179-207.
- Eaves, L. J. Testing models for variation in intelligence. *Heredity*, 1975, 34, 132-136.
- Eaves, L. J., Last, K., Martin, N. G., & Jinks, J. L. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *British Journal of Mathematical and Statistical Psychology*, 1977, 30, 1-42.
- Erlenmeyer-Kimling, L., & Jarvik, L. F. Genetics and intelligence: A review. *Science*, 1963, 142, 1477-1479.
- Eysenck, H. J. *Race, Intelligence, and Education*. London: Temple Smith, 1971.
- Gillie, O. Crucial data was faked by eminent psychologist. London *Sunday Times*, October 24, 1976.
- Havender, W. R. Sense and nonsense about the Jensenist Heresy. *The Alternative*, 1976, 9, 10-13. a
- Havender, W. R. Letter. *Science*, 1976, 194, 7. b
- Jensen, A. R. *Genetics and Education*. London: Methuen, 1972. a
- Jensen, A. R. Testimony to the Select Committee on Equal Educational Opportunity, United States Senate. In *Environment, Intelligence, and Scholastic Achievement*. Washington, D.C.: U. S. Government Printing Office, June 1972. Pp. 55-68. b
- Jensen, A. R. *Educability and Group Differences*. London: Methuen, 1973. a
- Jensen, A. R. *Educational Differences*. London: Methuen, 1973. b
- Jensen, A. R. Kinship correlations reported by Sir Cyril Burt. *Behavior Genetics*, 1974, 4, 1-28. a
- Jensen, A. R. How biased are culture-loaded tests? *Genetic Psychology Monographs*, 1974, 90, 185-244. b
- Jensen, A. R. Effects of race of examiner on the mental test scores of white and black pupils. *Journal of Educational Measurement*, 1974, 11, 1-14. c
- Jensen, A. R. Cumulative deficit: a testable hypothesis? *Developmental Psychology*, 1974, 10, 996-1019. d
- Jensen, A. R. The meaning of heritability in the behavioral sciences. *Educational Psychologist*, 1975, 11, 171-183. a

- Jensen, A. R. The price of inequality. *Oxford Review of Education*, 1975, 1, 59-71. b
- Jensen, A. R. Test bias and construct validity. *Phi Delta Kappan*, 1976, 58, 340-346. a
- Jensen, A. R. Equality and diversity in education. In N. F. Ashline, T. R. Pezzullo, & C. I. Norris (Eds.) *Education, Inequality, and National Policy*. Lexington, Mass.: Lexington Books, 1976. Pp. 125-136. b
- Jensen, A. R. Letter to the London *Times*, December 6, 1976. (Reprinted in: *Phi Delta Kappan*, 1977, 6, 471-492.) a
- Jensen, A. R. An examination of culture bias in the Wonderlic Personnel Test. *Intelligence*, 1977, 1, 51-64. b
- Jensen, A. R. Race and mental ability. In A. H. Halsey (Ed.) *Heredity and Environment*. London: Methuen, 1977. Pp. 215-262. c
- Jensen, A. R. Cumulative deficit in IQ of blacks in the rural South. *Developmental Psychology*, 1977, 3, 184-192. d
- Jensen, A. R. Genetic and behavioral effects of nonrandom mating. In R. T. Osborne, C. E. Noble, & N. Weyl. (Eds.) *Human Variation: Psychology of Age, Race, and Sex*. New York: Academic Press, 1978. a
- Jensen, A. R. The nature of intelligence and its relation to learning. In S. Murray-Smith (Ed.) *Melbourne Studies in Education*. University of Melbourne, 1978. b
- Jensen, A. R. Burt in perspective. *American Psychologist*, 1978, in press. c
- Jinks, J. L., & Fulker, D. W. Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, 1970, 73, 311- 349.
- Kamin, L. *The science and Politics of IQ*. Potomac, Md: Erlbaum, 1974.
- Lawrence, E. M. An investigation into the relation between intelligence and inheritance. *British Journal of Psychology. Monograph Supplement*, 1931, 16, No. 5.
- Loehlin, J. C., Lindzey, G., & Spuhler, J. N. *Race Differences in Intelligence*. San Francisco: W. H. Freeman, 1975.
- McAskie, M. Carelessness or fraud in Burt's kinship data: a critique of Jensen's analysis. *American Psychologist*, 1978, in press.
- McNemar, Q. Statistics can mislead. (Comment), *American Psychologist*, 1977, 32, 680-681.
- Morton, N. E. Analysis of family resemblance. I. Introduction. *American Journal of Human Genetics*, 1974, 26, 318-330.
- Munsinger, H. The adopted child's IQ: A critical review. *Psychological Bulletin*, 1975, 82, 623-659.
- Nichols, R. C. Black children adopted by white families. (Comment). *American Psychologist*, 1972. 32, 678-680.
- Plomin, R., De Fries, J. C., & Loehlin, J. C. Genotype-Environment interaction and correlation in the analysis of human behaviour. *Psychological Bulletin*, 1977, 84, 309-322.
- Scarr, S., & Weinberg, R. A. IQ test performance of black children adopted by white families. *American Psychologist*, 1976, 31, 726-739.
- Urbach, P. Progress and degeneration in the IQ debate. *British Journal of the Philosophy of Science*, 1974, 25, 99-135; 235-259.
- Van Valen, L. Brain size and intelligence in man. *American Journal of Physical Anthropology*, 1974, 40, 417-423.
- Viaud, G. *Intelligence: Its Evolution and Forms*. London: Hutchinson, 1960.
- Waller, J. H. Achievement and social mobility: Relationships among IQ scores, education, and occupation in two generations. *Social Biology*, 1971, 18, 252-259.