
Statistics and the Heritocracy

Angela Little

In 1912 Henry Goddard, a psychologist from Vineland, New Jersey, was invited by the United States Public Health Service to Ellis Island, where he applied a translation of the Binet Mental Test to arriving European immigrants. He reported that 79 per cent of the Italians, 80 per cent of the Hungarians, 83 per cent of the Jews, and 87 per cent of the Russians were "feeble minded". Five years later in 1917 the use of mental tests "for the detection of feeble minded aliens" had vastly increased the number of aliens deported.¹

Since their inception in 1905 when Binet and Simon were commissioned by the Minister of Public Instruction in Paris to design a test to identify children who were likely to fail in school, mental tests have arisen out of, legitimated, and reinforced aspects of the social order (e.g. income and race differences). The assumption behind many mental tests is that the personal dimension being tapped is of an innate, fixed nature and therefore largely unalterable under different social/environmental conditions. The development and modification of tests purporting to measure intelligence—defined variously as "an abstract reasoning ability" (Jensen 1969), "an innate general cognitive ability" (Burt 1921)—does involve the assumption that the full potential of any individual's development is limited by genic constraints laid down at the moment of conception.

The question of the genetic or environmental determination of human development has always been controversial. Psychologists and social commentators like Burt, Jensen and Herrnstein, argue that up to 85 per cent of individual differences in intelligence are accounted for by genetic factors. A more moderate estimate is made by Jencks. After a survey of all available data he concludes: "Our analysis of heritability suggests that the data base for making generalizations about the heritability of IQ scores is till too weak to justify a precise estimate. Our best guess is that genotype explains about 45 per cent of the variance in IQ scores" (1972). Others, like Borstelmann (1968), present an environmentalist position: that individual differences are explained by environmental factors.

The nature-nurture, genetic-environmental controversy is usually accompanied in the mind of the reader by the right-wing/left-wing political

construct—not surprisingly in many cases. Indeed, the actors at the centre of the debate, the persons who present the data, are not immune from political attitudes. As early as 1949, Pastore reviewed the writings of people concerned with the nature-nurture issue and found that the data presented, conclusions drawn and implications for society were heavily associated with their political and social attitudes. But the nature-nurture, right-wing/left-wing constructs do not completely overlap. There are many cases of persons who write allegedly 'conservative' pieces on the nature of society and whose assumptions about the individual are environmental. Durkheim's collective determination of the psychological concepts of time and space is one such example. On the other hand, Marx's dictum 'from each according to his ability, to each according to his needs' can be crudely interpreted as presenting a view of the individual who at birth has certain fixed, fairly unalterable abilities.

What I propose to do in this paper is to examine the inferences which can and cannot be drawn from the statistics presented in the debate. Secondly, I wish to draw attention briefly to elements of arguments about intelligence in society which often remain submerged, and finally, to suggest a few implications which IQ statistics have for developing countries.

The array of data surrounding the nature-nurture controversy is vast. I have already referred to the two key concepts in the debate, 'heritability' and 'variance'. Variance refers to the total amount of variation of scores on IQ tests among a given population. This total variance is regarded by Jensen and population statisticians as the sum of three main components, each contributing their own amount of variance: genetic factors, environmental factors and error factors. The proportion of the total variance due to heredity factors is referred to as the 'heritability coefficient', or 'heritability'.

How then is a heritability coefficient computed? For a large number of pairs of individuals (always of the same kinship relation) the test results of one of the pair are correlated with the results of the others. The kinship relation usually regarded as the most reliable source of data is the monozygous twin relationship. (Two persons developed from a split zygote (egg), and therefore in theory having the same gene structure or genotype.)

¹ Historical evidence quoted in Kamin, L. (1973).

Where monozygous twins have been reared apart, the proportion of variance in test scores due to environmental factors can be computed.

Once the heritability coefficient is derived though, the figure is often given a number of erroneous meanings. The statistical concept of variance defined above refers to the total amount of variation between individuals, or, to be more specific in this case, the sum of the differences between the scores of each person in a pair of monozygotic twins. We are looking at differences **between** individuals. The famous heritability figure of 80 per cent is often misinterpreted to mean the proportion of any given individual's total development which can be accounted for by genetic factors. This is nonsense. Heritability is a population statistic. It operates at a level beyond the individual. Jensen makes this point clear:

"heritability . . . has no sensible meaning with reference to a measurement or characteristic in an individual. A single measurement by definition, has no variance. There is no way of partitioning a given individual's IQ into heredity and environmental components . . ."

(Jensen 1969 p. 42).

Now it may be possible to make probability statements about the proportion of a difference between two individuals attributable to genetic or environmental factors, but it is impossible to take one individual and to attribute a certain proportion of his total development to genetic or environmental factors, even on a probability basis. It is theoretically possible that a population of individuals might yield a heritability value of 90 per cent whereas the actual genetic proportion of any given individual's total development (not that we could ever measure it) might only be 10 per cent. Heritability refers only to the differences between individuals, and such differences may actually represent a very small proportion of an individual's total development.

We can take as an analogy the concept of group cohesion, which refers to the degree of 'togetherness' displayed by a group of persons. A group in which the individual members displayed little mutual attraction would be lacking in group cohesion. Now an index of group cohesion might allow us to state the probable degree of mutual attraction between two members of the group, but it would be meaningless to use this index to make a statement about the internal cohesion or integrity of an individual. Again this would be impossible, even on a probability basis. Indeed, a group may well be cohesive precisely because the people who comprise it are **lacking** individual integrity. In short, heritability is a statistic beyond the level of the individual and it cannot be used

to support a statement about the proportion of an individual's total development attributable to genetic or environmental factors.

A second characteristic of heritability is its dependence on the sampled population. Heritability is an empirically determined population statistic, and like any statistic its value is affected by the characteristics of that population. Heritability will be higher in a population in which environmental variation relevant to the trait in question is small than in a population in which there is great environmental variation. This raises questions about the limits of a population and the generalizability of the twin samples to a population. If we are to apply the dictum that heritability coefficients apply only to the population sampled, then it is tempting to conclude that coefficients derived from separated twin studies apply only to the total population of separated twins. This may be taking the point too far but the separated twin studies are most certainly not representative of, for example, the black population. The significance of this is realized when we see how Jensen uses a heritability estimate, from a North American white sample to explain the 15-point inter-group difference between the average IQ scores of a large white population and a large black population. To my knowledge a heritability estimate has never been calculated for a sample representing the total black North American population. Suppose for example, a heritability coefficient were computed and turned out to be as low as 20 per cent for the black population. How then would we explain the 15-point difference? Why should we use the coefficient of 80 per cent rather than the other coefficient of 20 per cent to explain the difference? The only way to compute the proportion of variance between blacks and whites attributable to genetic factors would be to include a study of white monozygotic twins raised in separate black environments, and a study of pairs of one white/one black monozygotic twins. The former situation is unlikely, given present social constraints, and the latter is a genetic impossibility.

Not only are estimates of heritability specific to the population sampled but they are also specific to the particular tests used, the method of measurement and the particular point in time. Jensen himself presents data to show how heritability estimates depend on the intelligence test used. Estimates varied from 0.42, when a Swedish IQ test was used, to 0.80 when the Stanford Binet test was used.² The relevance of the method of

² While Jensen demonstrates that heritability depends on the test used, Jencks shows how the type of heritability coefficient used determines the result. When different heritability formulae are applied to the same data the coefficients may vary from 0.60 to 0.91. (Jencks 1972 p. 294).

measurement has been revealed in several studies which demonstrate the interpersonal and expectancy effects of the testing situation—white testers record scores from black subjects significantly lower than those obtained when a black person does the testing; results from twins are closer when tested by the same psychologist than when each is tested by a different psychologist. The dependence of a heritability coefficient on a point in time reminds us of the heritability estimates for behaviour characteristics in animals, which have been found to increase, decrease and fluctuate randomly as a function of repeated testing.³ Although IQ tests have themselves been tested repeatedly on humans in order to select out items that are inconsistent over time, I do not know of any study which actually makes repeated heritability estimates on the same population over fairly small intervals (say one day). So much, then, for the reliability of heritability.

My second set of points is about how IQ data relate to other social indicators within a society. Studies showing the positive intercorrelation of intelligence test scores, educational achievement and social economic status, are numerous (Blau and Duncan 1967, Jencks 1972). Scores are often grouped together by socio-economic class and we see that the highest status group has on average the highest IQ score.

Such data are often used to support political arguments: those of higher status are the most capable members of society and therefore deserve their rewards; society would be run more efficiently if we ensured that all those obtaining high IQ scores were allocated to high status positions; there are certain jobs in society which require scarce mental abilities etc. The use of such data in these arguments overlooks a number of aspects of the relationship between data on IQ and social class.

The range of scores within a class is often greater than the differences between classes. Intelligence tests usually favour the group on which the test is standardized initially (in the case of the most widely used test, the Stanford Binet, this was a white middle-class sample). The most frequently used validation criterion for intelligence tests is school success, i.e. if we want to know whether a test measures what we intended it to measure, we compare a group of individuals results with their results on another measure. When intelligence tests were first developed this second measure was school success. Now it is usually another intelligence test which itself was validated

against school success at some point. Therefore, it is not surprising to find that IQ correlates with school success which itself correlates with job success and socio-economic background.

These political arguments advance a 'realized potential' concept of intelligence. The higher social strata comprise those who, on average, have higher IQ scores. Certain occupational roles tend to be performed by people in these strata. Therefore, certain minimum levels of intelligence are necessary to perform these jobs. But this argument is valid only if we assume that everyone develops to his full potential at some point in his life and continues to express it. Only if this is true can we argue that those performing occupations of lower status and having lower IQs are incapable of performing higher status occupations, and performing them **just as efficiently**.

Now this sort of assumption is difficult to prove either way. However, we can depend on human intuition to refute it and to remind us of the scores of schoolmates whose talents appear to have been wasted as they dropped into one of society's slots; the creativity of very young children's thoughts and actions which convention deems it necessary to curb; the wealth of thoughts and actions of the older generation which were perhaps once expressed but are now inhibited by Western society's expectation that old people make only negative contributions. It is unlikely that many people realize their full potential. The majority of people operate below their maximum. Ultimately, genetic constraints may become apparent, but the overwhelming and more immediate constraints in the present world are man-made—the constraints of political, social and economic systems. The 'realized potential' concept of intelligence provides its own proof—people (and perhaps societies) who interpret failure in terms of their own inherent lack of ability are unlikely to invest effort in improving themselves. While it is unlikely that the majority of people ever realize their full potential, a social concept of intelligence which implies further constraints ensures that they never will.

A step towards testing the assumption about realized potential might be made if certain social institutions were prepared to undertake social experiments. For example, an employing organization could undertake (perhaps with a risk-taking monetary guarantee for both employer and employee) an experiment in which a number of persons of widely differing intelligence levels perform the same job. Universities might evaluate the performance of students admitted on a lottery basis, irrespective of educational background. (While University teaching

³ Studies in animals which show the dependence of heritability on a point in time, are quoted in J. Hirsch (1970). This article also provides further comment on the misuse of the heritability statistic.

staff might evaluate the students, the admissions policy would need to be concealed lest their academic prejudices colour their perceptions and grading of students.)

Some implications for developing countries of IQ statistics should now be becoming clear. It is sometimes argued that certain countries are 'under-developed' because their populations are of lower intellect than those of the developed world. The points elucidated earlier about heritability estimates and the differences between racial groups cannot be overstated. It is possible to demonstrate that one racial group in society performs at an inferior level on an IQ test, but we must remember that the group which devises and administers the test invariably comes out on top. It is also possible to correlate differences between societies or racial groups on some measure of economic development (e.g. income distribution) with differences on a measure of intellectual development (e.g. average IQ score). We can even make the bold assumption that a causal arrow runs from intellectual development to economic development, and attempts to explain economic differences in terms of intellectual differences. But we will never be able to explain the intellectual differences in terms of either genetic or environmental factors. To explain differences between societies requires a number of activities—the co-operative development of an intelligence test which would reveal differences between the average performances of the groups who designed it, and the computation of a heritability estimate which would apply across countries or racial groups. As explained above, in the case of blacks and whites this would involve a number of impractical combinations of environment and genotype.

Undoubtedly there are genetic upper limits to the activities of human beings and societies, but it is unlikely that such limits will ever be reached. To blame an individual or society's shortcomings on 'inherent' lack of ability is an easy way out. It is always more difficult, but probably more honest, to accept that such shortcomings are man-made rather than made in man.

A second set of implications of IQ statistics concerns their use as a social technology. Many developing countries are in the process of developing local IQ and aptitude tests for a variety of purposes, e.g. educational and occupational selection. The development of local IQ tests sometimes means simply changing the content of American or British tests to fit local conditions.

Alternatively, tests may be built up from scratch by indigenous or Western psychologists. But in such cases the **process** of test construction—item

selection, test validation, standardization etc.—is the same as in the West. The influence of the US educational testing service is very apparent in many developing countries.

Now there is nothing wrong with a developing country deciding to import a process which has been developed in the West (even if it does contain implicit assumptions): what is wrong is to import a process which does not explicitly state its assumptions. One such assumption is that the distribution of intelligence follows a normal curve. Many psychologists in the West, forgetful of their own assumptions, write as though IQ scores are actually distributed normally, but in fact the normal distribution of IQ scores is nothing more than a statistical artefact: items are selected and rejected according to whether or not the total scores for a large population follow a normal distribution. On this point Vernon (1960) has written: "it is practically impossible to prove that mental abilities are normally distributed. But there is no logical reason, nor any strong evidence, for suggesting any other type of distribution. Psychologists have accepted this dogma since it provides a convenient basis for test construction and for statistical analysis of test scores". Another related assumption in the use and construction of IQ tests is that the standard deviation of IQ scores on a good test should be 15 points either side of the mean score of 100. The standard deviation refers to the range of scores which accounts for approximately 68 per cent of the population of IQ scores. When the standard deviation is 15 points, 68 per cent of the scores should fall within the range 85 to 115 points. Now there seems to be no obvious reason why a population of IQ scores should not have a standard deviation of 20 points of 5 points instead of 15. All this means is that the shape of the distribution would be more or less flattened. The flatness of a distribution will be determined by the degree to which testers choose to discriminate between people.⁴

Intelligence testing as a social technology provides a useful tool for allocating persons to jobs. It also adds 'scientific fact' to the legitimation of any process which has to allocate people to slots in an unequal economic and social structure. The difference between allocation on the basis of criteria like class or race on the one hand, and 'intelligence' on the other is that intelligence, given scientific credence, may be more resistant to change in the future than class or race have been in the past.

⁴ For comprehensive accounts of the construction of intelligence tests, see Ohuche and Akeju (1975), Ryan (1973) and Simon (1971).

Bibliography

Jensen, A. R., 1969, 'How much can we boast IQ and scholastic achievement', **Harvard Educational Review**, Vol. 39 No. 1, 1-123.

Burt, C., 1921, **Mental and Scholastic Tests**, Staples Press, London.

Jencks, C., 1972, **Inequality**, Basic Books, New York.

Borstelmann, I.J., 1968, 'The Culturally Disadvantaged and Compensatory Education: Fantasies and Realities,' **Newsletter** No. 9 of the Association of Educational Psychologists.

Pastore, N., 1949, **The Nature-Nurture Controversy**, Geoffrey Camberlege, London.

Hirch, J., 1970, 'Heritability and Racial Intelligence, Simplism and Fallacy: irresponsible pollution in channels of intellectual communication,' Cambridge Society for Social Responsibility in

Science: a Public Debate on Education, Race, IQ and Intelligence. July 18, 1970.

Vernon, P. E., 1960, **Intelligence and Attainment Tests**, ULP, London.

Kamin, L. J., 1973, Heredity, Intelligence, Politics and Psychology. Princeton University, mimeo.

Blau, P. M., and Duncan, O.D., 1967, **The American Occupational Structure**, Wiley, New York.

Herrnstein, R. J., 1973., **IQ in the Meritocracy**, Allen Lane, London.

Simon, B., 1971, **Intelligence, Psychology and Education**, Lawrence and Wishart, London.

Ryan, J., 1973, 'The Illusion of Objectivity', in K. Richardson and D. Spears (eds.), **Race, Culture and Intelligence**, Penguin, Harmondsworth.

Ohuche, R. O. and Akeju, S. A. 1975, **Testing and Evaluation in Education**, African Educational Resources, Monrovia.