WHAT IS INTELLIGENCE?
A Re-examination

Douglas A. Samuelson

Intelligence is like charm or good looks: we all know what we mean by it, we all agree that it is good to have, but we cannot define just what it is or where it comes from. Most of us think of it as something like "general ability to solve complex or obscure mental problems," perhaps emphasizing abstract reasoning and pattern recognition.

Like "random" or "probability," though, the word holds pitfalls for the unwary. The more one knows about it, the more difficult it becomes to define. The Encyclopedia of Education (1971) states that no definition is universally accepted, then devotes several paragraphs to a discussion of what a good definition should include.[1]

The history of the concept is both enlightening and surprising. In 1905, Alfred Binet and his pupil T. Simon developed the first of what are now known as intelligence tests in response to a commission from the Paris public schools soliciting ways to distinguish "feeble-minded" from "normal" children so that the former could receive special instruction. The term "intelligence," derived from a Latin root meaning "to choose," was adopted by Binet and Simon to describe the mental attribute measured by their test: judgment, rather than sensory perception or memory. Indeed, the idea that "intelligence" is best thought of as one quantity, rather than a set of not-necessarily-related attributes, gained acceptance mainly because Binet's attempts using the single-valued tests succeeded after "cluster" and "profile" tests had failed.

What followed Binet's pioneering work is familiar to most students of psychology. Lewis Terman of Stanford adapted and improved the test for American children, and the test he devised, through its continuing evolution, has become the standard for the measurement of intelligence. The widespread use of the Stanford-Binet test to determine the mental abilities of World War I draftees popularized the test and the concepts behind it, and by the late 1920's its use in the schools was commonplace, together with many of the doubts, debates and abuses still common today.

However, the important constant in this history is that tests of intelligence have invariably been validated by their ability to predict success in the traditional Euro-American school system. As psychologist Arthur Jensen took care to note in his comprehensive and controversial article, "How Much Can We Boost IQ and Scholastic Achievement?"[2]:

> "The content and methods of instruction represented in this tradition, it should be remembered, are a rather narrow and select sample of all the various forms of human learning and of the ways of imparting knowledge and skills....We have accepted traditional instruction so completely that it is extremely difficult even to imagine, much less to put into practice, any radically different forms that the education of children could take. Our thinking almost always takes for granted such features as beginning formal instruction at the same age for all children (universally between the ages five and six), instruction of children in groups, keeping the same groups together in lock step fashion through the first several years of schooling, and an active-passive, showing-seeing, telling-listening relationship between teacher and pupils."
>
> The consequences of this history were clear: "Satisfactory learning occurs under these conditions only when children come to school with certain prerequisite abilities and skills: an attention span long enough to encompass the teacher's utterances and demonstrations, the ability voluntarily to focus one's attention where it is called for, the ability to comprehend verbal utterances and to grasp relationships between things and their symbolic representations, the ability to inhibit large-muscle activity and engage in covert "mental" activity, to repeat instruction to oneself, to persist in a task until a self-determined standard is attained--in short, the ability to engage in what might be called self-instructional activities, without which group instruction alone remains ineffectual."

Therefore, the characteristics which Binet and Simon found most useful in distinguishing "bright" from "dull" children, and all the testing items and procedures which have descended from their efforts, reflect the value judgment that, as Jensen put it, "despite all the criticisms that can easily be leveled at the educational system, the traditional forms of instruction have actually worked quite well for the majority of children. And the tests that were specifically devised to distinguish those children least apt to succeed in this system have also proved to do their job quite well."[3] (Emphasis mine.)

Thus the relationship between tested intelligence and intelligence as socially defined depends upon the assumption that the skills most important to scholastic success are highly related to what we think of as "general mental ability." This is not an unreasonable assumption, since the educational systems have been directed toward teaching and improving the skills considered most important in later life; also, school success and standardized tests are major screening devices used in determining who will have access to the best opportunities later, which reinforces the apparent predictive value of intelligence tests.

However, recognizing this assumption expli-

citly has several important consequences. Let us examine the statistical consequences first.

IQ As a Statistic

As we have seen, IQ and other related single-valued measures of intelligence were developed from observation of behavior of "bright" and "dull" school children, and the tests were and are validated by prediction of scholastic success. Although Binet and Simon could not have known it, since the relevant statistical theory was not invented until many years later, their attempts to summarize the children's distinguishing traits in a single score can be viewed as a rather primitive, ad hoc sort of discriminant function analysis. They selected types of behavior, assigned weights to scores on items which measured different types of behavior (the weights presumably were selected so as to maximize the difference in average scores between "bright" and "dull" children), and then formed the weighted sum of these item scores.

Since this approach required the use of a wide variety of unrelated items varying over a continuum of difficulty, it should not surprise statisticians that the weighted sum had a roughly normal distribution. (In fact, Jensen's subsequent analysis suggested that the accumulated U.S. data can be fitted even better with a double-normal distribution, with one normal "hump" at mean of 55 to 60 for children with severe retardation caused by known genetic or nutritional defects, and a much larger normal "hump" with mean about 105 for all other children.[4] This would fit the proposed description of IQ as a discriminant function even better.)

Having achieved success with their method, Binet and Simon had little explanation for the underlying structures or attributes; but others were quick to supply ideas. Their results were published at the height of the "Social Darwinist" movement in American and European philosophy, as various thinkers -- notably Herbert Spencer -- asserted that man, like the animal species, is undergoing natural selection which manifests itself through the social class structure.

The approximately normal distrubution of IQ scores led nicely into the hypothesis that intelligence was a polygenic trait like height; the leading analysts who tackled the subject-- Francis Galton, Lewis Terman, Cyril Burt, Henry Goddard and Robert Yerkes, among others -- were all hereditarian and "Social Darwinist" in their thinking; several of these thinkers also belonged to one or more "eugenicist" societies, advocating breeding of humans as a means to improve the species and solve social problems.[5]

So soon after the idea was first developed, theory about intelligence was pushed into a hereditarian mold which has shaped all subsequent discussion. The discriminant function procedure, with its great robustness against violation of assumptions, would naturally tend to prevent subsequent data from providing a striking contradiction of theory -- or any other theory that had been adopted, for that matter.

When psychologists attempted to extend intelligence testing to adults, they soon found that the "mental age divided by chronological age" definition was meaningless, since the raw scores of adults of 50 were approximately the same as those of adults of 25. "Mental age" did not make sense for adults. Therefore, using the hereditarian theory which had since been developed, they simply assumed that IQ was normally distributed and standardized their test scores accordingly: that is, they assigned an IQ score of 100 to the mean raw score, 85 to the raw score attained or exceeded by 86% of the sample popluation, 115 to the raw score attained or exceeded by 14% of the population, and so forth.

All modern intelligence tests use this procedure, the last holdout -- the Stanford-Binet -- having converted to it in 1960.[6] Thus the normal distribution of IQ scores is a direct consequence of the scoring procedure itself; this distribution can no longer be said to prove anything about genetic influences.

In summary, then, it seems most reasonable to conclude that tested intelligence -- IQ and related measures -- can best be regarded as a statistical estimate of the probability of scholastic success, derived from a form of discriminant function analysis. This statistic is robust against the vagueness of the underlying theory and is insensitive to changes in critical assumptions, and may therefore understate or camouflage effects of theoretical errors. To illustrate the potential magnitude of this problem, let us consider a hypothetical example.

Viewing IQ From Afar: A Cautionary Tale

Imagine a school serving two neighborhoods, one primarily English-speaking, the other primarily Spanish-speaking. Suppose that instruction in the school is in English only, and that the two groups of children have equal means and distributions of IQ if tested in their own languages. Now suppose that a psychologist downtown at the school board office, ignorant of the situation, requests that a standard English IQ test be given to all the students in the school.

First of all, he would see that the scores for English-speaking students correlate well with teachers' opinions of them and with other measures of achievement and ability. The Spanish-speaking group would have depressed scores, with those who knew the most English doing best; this, too, would correlate highly with scholastic success. Finally, he might be tempted to conclude that the Spanish-speaking group was seriously inferior in general mental ability.

Naturally, the school counselor would soon propose a program to remedy the difficulty: tutoring the Spanish-speaking students in English. New IQ tests after several months of this program would show substantial gains for the Spanish-speaking students, and their school performance would also improve.

But what would happen after the tutoring

ended? Retests a year or two later would show the Spanish-speaking students' IQ scores and school performance both declining again. The school board psychologist would say, sympathetically, "Well, it seemed like a good idea, but all you got was a 'hothouse effect.' As soon as you stopped the intensive tutoring, they started sinking back to their 'natural' level." And we can imagine him writing to his superiors, as Jensen did in his 1969 article,[7] "Compensatory education has been tried and it apparently has failed."

Now the frustrated counselor might try another approach: he translates the IQ test into Spanish and gives another school-wide test, with each child being tested in his own language. Would this change the situation? No, the school board psychologist, reviewing the new score distributions, would notice that the new test has a lower correlation with school success, and would therefore pronounce it "invalid."

This is why the validation of intelligence scores by prediction of school success is so important. If we have overlooked a major factor or group of factors which determine IQ, we may well be led to an error just a great, though not as obvious, as that of the hypothetical psychologist.

For example, for "English-speaking" and "Spanish-speaking" substitute "white" and "black"; for language, substitute tendency (perhaps largely or wholly genetic in origin) to react adversely under stress. Is it not possible that much of the observed difference in scores of black and white U.S. school children could be attributable to such factors? Even more important, given the way in which intelligence is evaluated and interpreted, how would we know whether such factors were involved? With these questions in mind, we turn to the core of the recent controversy: heritability.

Intelligence and Heritability

The concept of heritability, particularly as applied to race difference in intelligence, has been exhaustively reviewed and discussed; it has been the focus of the controversy over racial differences in intelligence. To summarize briefly, heritability has been estimated mainly from studies of twins reared apart, and of unrelated foster children reared together. Correlations between IQ scores of identical twins reared apart have averaged around .8; correlations between unrelated children reared together have averaged around .5 to .6. This evidence and corroborative estimates derived from the differences between correlations for monozygotic (identical) and dizygotic (fraternal) twins have led to a tentative conclusion, widely but not universally accepted, that the most reasonable estimate for heritability of IQ is 60 to 80 percent.[8]

Even if we concede, for the moment, that heritability is very important and high heritability is very discouraging to "environmentalists" -- a point to which we shall return later -- two important criticisms must be made. First, heritability estimates of IQ mis-attribute some environmental factors to heredity because of the structure of the experiments. Second, heritability estimates of IQ are artificially inflated because of the way the tests are constructed and updated.

The first point comes from the fact that heritability involves not two factors, but four: genotype; environment; covariance (the fact that genotypes may not be assigned to environments at random); and interaction (the fact that different genotypes may respond differently to a given environment). While some critics have raised questions concerning the possible neglect or understatement of the latter two factors,[9] most seem to have missed the fact that neither covariance nor interaction can be estimated from twin and foster-pair studies, and both effects will show up as "hereditary" in such studies.

To see why this is so, consider a hypothetical study in which two white twins, say Jim and John, and two black twins, say Bob and Bill, are separated at birth: Bob is reared with Jim and his white parents, John is reared with Bill and his black parents. Jim and Bob are considered by the methodology of twin studies to have the same environment, and accordingly all differences between their test scores would be attributed to heredity; in the same way, differences between John's and Bob's scores would be attributed to heredity.

But these differences include the different treatment Jim and Bob receive in school and elsewhere because one is white and the other is black. The same is true for John and Bill. Thus <u>the effects of direct racial discrimination</u> (interaction effects) <u>are attributed to heredity</u>.

Past studies and standardizations have suggested that interaction effects are small -- on predominantly white samples. A true "crossing experiment" such as the one just described has never been done, and probably never could be done within the code of ethics governing experimentation on human subjects.

In this example, covariance between heredity and environment would be zero. Similarly, though it is clear that covariance, like interaction, cannot be estimated in a twin or foster-pair study, and the effects are most likely to be attributed to heredity. Actual adoptions are possible only for foster parents who are willing to adopt and who meet certain criteria set by the regulatory agencies. It cannot reasonably be argued that assignment of children to foster parents is anything resembling random. The tendency one would expect is, first, that the range of environments available would be substantially narrower than the range presented to natural children; and, second, that the white children would have a high probability of getting better environments. Again, similarities between white twins reared apart -- including the better-than-random assignment to environments -- would be attributed to heredity, and similarly for the effects of the non-random assignment of black children to environments. In addition, the limited range of environmental variation would further inflate

heritability.

The second point, which no one to my knowlege has noted, is that the dependent variable (intelligence) changes over time: new test items are continually being developed, old items have been discarded as "no longer meaningful," and the rules of scoring and interpretation have changed accordingly. Intelligence, unlike, say, height, cannot be said with assurance to be the same thing in 1970 as in 1920. To understand the importance of this criticism, we must remember that items are dropped from IQ tests because they no longer predict well, and new items are added because they predict better. Thus some of the most important effects of environmental changes are attributed in part to obsolescence of test items.

For example, tests which included such "general knowledge" items as the identity of New York's baseball team were used in the 1920's to support the conclusion that recent immigrants from Eastern and Southern Europe were "inferior" to Americans of Western European ancestry.[10] Such items no longer predict success in school effectively, so they have been dropped and replaced; the conclusion drawn from the tests at the time, that Eastern and Southern European immigrants were "genetically inferior," has also been quietly discarded; but the heritability estimates derived from the tests remain in the scientific literature.

The IQ's of current descendants of those "inferior" immigrants are much higher than the high heritability estimates of that time would have led us to expect, but instead of being viewed as strong "environmentalist" evidence, these results are attributed to obsolescence or cultural bias of the tests items used then, if they are discussed at all.

These and other technical criticisms cast considerable doubt upon the value of heritability estimates in drawing scientific conclusions and in determining policy. Overshadowing all these arguments, however, there remains the question of what heritability of intelligence really means.

Is Heritability Important?

As we have seen, the usefulness of heritability in thinking about intelligence is diminished considerably by the technical difficulties in measuring heritability in a statistically sound manner. Even if heritability could be measured in a manner which would answer these objections, though, it is not clear what we would gain.

No matter how it is measured, heritability reflects only the effects of the environments and genotypes which were included in the sample from which heritability was estimated: for example, tuberculosis was highly heritable before the discovery of modern antibiotics, when incidence and severity of TB depended mainly on such largely genetic traits as respiratory allergies and general susceptibility to infection. In the same vein, it is possible to construct an experiment in which two groups of plants are selected from the same set of genotypes; one group is raised in light, the other in darkness. The heights of the plants will be highly heritable within groups, but the difference between groups will be environmental.

Similarly, the discovery of new environmental treatments which could alter IQ would also lower the heritability of IQ. High heritability does not "prove" that the trait cannot be modified; it suggests that modification through treatments previously tried is unlikely. It is unwarranted to concede, as many critics of the "Jensenist" position have implicitly done, that defensibly high heritability estimates for IQ would demolish the justification for trying to modify IQ through research and social programs. At best, heritability is useful primarily as an indication of which types of treatment seem unlikely to succeed. If one is tempted to agree with Jensen that methods of compensatory education hitherto tried have been proven unsuccessful, the conclusion should be to try altering the factors which have remained unaltered: starting age for school, home environment, prenatal nutrition, and so on. But perhaps the most important unchallenged factor is the assumption that the "intelligence tests" which predict success in school are actually measuring the general mental abilities which are needed for success in life.

Implications for Policy and Research

If the current controversy is seen, then, as much fuss over the wrong questions, where do we go from here? The first step, I think, is research on the right question: what do intelligence tests really measure? If tested intelligence is really a sort of discriminant function, why not see whether a properly designed discriminant function can't do better? Specifically, it should be possible to perform a discriminant function analysis on children classed as "bright" or "dull" according to school performance, using demographic factors and various test results as independent variables. If all the predictive information could be supplied by demographic variables, would that not change our ideas about the determinants of school success? Whether IQ provides independent information or not, would it not be useful and informative to learn what else is important in predicting school success?

Continuing research into motivational psychology and various observation-based studies of developing perception and reasoning -- the work of Piaget and his followers, for example -- is surely more promising than further explorations of the "nature-nurture" controversy. As was suggested earlier, experiments which could resolve the latter problem are almost certainly impossible under the code of ethics governing human experimentation; moreover, even if the experiments could be performed, their results might be rendered irrelevant by the discovery of a new treatment or by the acceptance of the idea that intelligence tests aren't really important anyway.

For policy, these conclusions are clear:

-- Remedial-intervention programs should

concentrate directly on skills, not on raising IQ. If school performance changes, IQ scores will have to change in order for the tests to remain acceptable as predictors.

-- Research or policy based on "eugenic" considerations should be viewed with great suspicion until the meaning of tested intelligence is clearly understood.

-- Perhaps most important, before proceeding with massive programs to search for treatments to raise intelligence, we should consider not only whether raising intelligence is useful -- which, I suggest, is an unanswerable question until we understand clearly what intelligence is -- but also how the fair administration of the treatment could be controlled. If intelligence were largely hereditary, at least the resultant differences in education and occupational status could be rationalized sufficiently for society to function. But what if a simple environmental treatment -- say, a "smart pill" or hormone injection -- were found? Who would control who gets how much? And how would the "less fortunate" react?

Perhaps the most useful consequence of the race-IQ controversy is that social scientists and the public may be forced to realize that science cannot be divorced from its social context, as assumptions alter conclusions. The next time some data (such as the IQ scores) suggests to some analysts that the democratic ideal is unworkable, we should re-examine the analysis and the data before preparing to discard the ideal. Both science and democracy will be the better for it.

References

1. Encyclopedia of Education, (MacMillan, 1971), pp. 672-3.
2. Harvard Educational Review, Winter 1969, p. 7.
3. Ibid., p. 7.
4. Ibid., pp. 23-7.
5. This idea is developed, perhaps with more stridency than is warranted, in Leon Kamin's book, The Science and Politics of IQ (Wiley, 1974). Other useful references on the history of the concept include "Social Implications of IQ," by Sheldon H. White, in National Elementary Principal, March/April 1975 (there is also a short article by Kamin in the same issue), and Chapters 1 and 2 of Race and Intelligence, Ken Richardson and David Spears, editors, (Pelican, 1972).
6. Encyclopedia of Education, op. cit., pp. 131-2
7. Harvard Educational Review, op. cit., p. 2.
8. A well-written and fairly comprehensive review is Chapter 4 of Loehlin et. al., Race Differences in Intelligence (W. H. Freeman and Company, 1975).
9. Chapter 4 of Loehlin et. al. (op. cit.) briefly reviews the criticisms raised by Jencks (Inequality: Basic Books, 1972), Kamin (op. cit.), and Layzer (Science, 1974, 183, 1259-1266). The spring, 1969 issue of the Harvard Educational Review included an article by R. Light and P. Smith which was the first to introduce interaction into the discussion, although not very persuasively. W. Bodmer and L. Cavalli-Sforza were apparently the first to argue that the "nature-nurture" controversy can never be resolved under present circumstances because of confounding of interaction, covariance, and heredity. (Scientific American, October 1970, pp. 19-29.)
10. Kamin (op. cit.) discusses this at length.