Die Naturwissenschaften 68. Jahrgang Heft 10 Oktober 1981

The Structure and Measurement of Intelligence

H.J. Eysenck
Institute of Psychiatry, University of London

One of psychology's outstanding successes has been the measurement of intelligence, and the demonstration that differences in intelligence, so measured, were due in large part to genetic factors. In recent years much work has been done to clarify the problem of the biological basis of these inherited differences, and work on the evoked potential in the EEG has generated important new findings in this field. We now know far more about intelligence, its inheritance, and its biological basis than we did even a few years ago.

Science and Intelligence

Scientists are often doubtful about the possibility of measuring what to them seem quite insubstantial psychological concepts, like intelligence, personality, etc. In principle such objections do not have any philosophical substance; all scientific concepts are necessarily insubstantial. What applies to intelligence applies equally to heat, gravitation, or mass. All are concepts which need to be embedded in a theoretical framework in order to make meaningful measurement possible. In my book on "The Structure and Measurement of Intelligence" [1] I have made a careful comparison between the measurement of intelligence and the measurement of heat, trying to show that there is a very close analogy between the two. Both start with everyday observations – in the one case of sensations of heat and cold attending the touching of snow and ice, or exposure to the sun, the other with experiences of problem solving which can be easy and quick, or slow and laborious. Both go on to construct theories, make measurements based on the theories, and gradually, in a corkscrew fashion improve both theory and measurement. In both cases there are still anomalies to be ironed out. Thus in the measurement of heat, different methods of measuring temperature, such as the mercury-in-glass thermometer, the constant-volume thermometer, the resistance thermometer, and the thermocouple give different readings for identical conditions – thus when a mercury-in-glass thermometer reads 300 °C, a platinum-resistance thermometer in the same place and at the same time will read 291 °C! On the theoretical side, too, there are essentially two different theories of heat, the thermodynamic and the kinetic, alerting us to the fact that it would not necessarily be reasonable to expect psychologists to have worked out a unitary theory of intelligence, accepted by all. If we had to wait until such a theory existed before declaring a subject "scientific", and "measurable", then we would ban, not only heat, but also gravitation – there we have the action at a distance theory of Newton, the field theory of Einstein, and the quantum mechanic theory of particle interaction, all useful in their way, but still falling short of a single, unified theory.

Any theory of intelligence is based on observations already made by the ancient Greeks, separating out cognitive from emotional aspects of behaviour, i.e. those aspects which involve ideation, problem solving and thinking from those which involve emotional reactions, motivation, and will. To the former Cicero gave the name "intelligentia", and this term began to be used in a technical sense by Sir Francis Galton, a cousin of Darwin's, towards the latter half of the last century. When speed and accuracy of cognition are measured in relation to widely different types of problems, it is generally found that there exists what is technically known as a "positive manifold". In other words, when hundreds or thousands of people are submitted to these tests, and the results are intercorrelated, then it is found that all the correlations are positive – in other words, a person who is good at one type of intellectual process tends to be good at all others. Note the use of the word "tends"; the correlations are not perfect, and therefore we are dealing with statistical relationships which are not invariant. Nevertheless, these statistical relationships are stable from time to time and from place to place; they obtain equally in many different countries and cultures as they do in the Western World.

Tests of Intelligence

What is the nature of the tests used in these determinations? Psychologists make a careful distinction between intelligence and acquired knowledge, and try to measure the former by using test materials which are all equally known or equally unknown to all the persons who are being tested. The distinction is of course an obvious one; I can solve many mathematical problems that would have been insoluble for Newton or Euler, not because I possess a greater amount of intelligence, but because I have acquired a certain amount of knowledge that was not available at the time these two men were working and writing. Typical of an intelligence test item which does not rely on acquired knowledge is the matrix type test shown in Figure 1, where the top of the diagram shows the 3×3 matrix of geometric figures, including a blank; the correct answer has to be chosen from the six figures given at the bottom, and the number of that figure has to be written into the empty space. The problem contains all the elements necessary for its solution; no prior knowledge is required (except

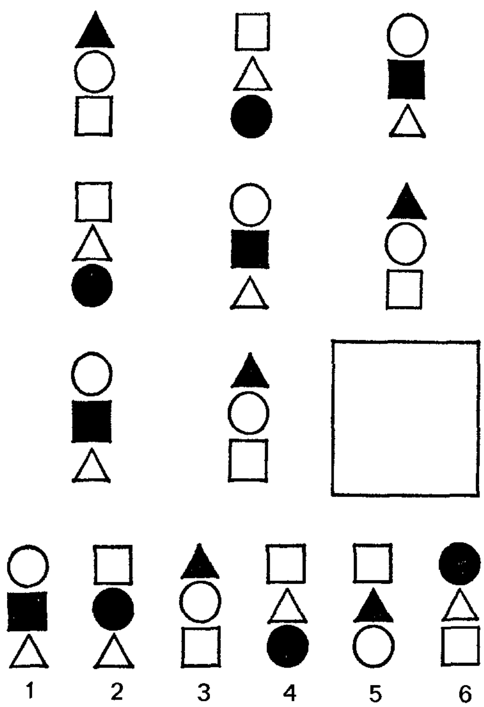

Fig. 1. Typical non-verbal, culture-fair intelligence problem

of course very obvious items like knowing how to hold a pencil, how to make a mark on the paper, and how to read numbers from 1 to 6!). Tests of this kind can be made very easy or very difficult, and they have been found to be excellent measures of intelligence.

What is involved in a tests of this kind is essentially the induction of relations between different parts of the test, and, based on that, the induction of a correlate. Thus in each row, and in each column, we have a triangle, a circle, and a square. Some of these are black, others are white, in each row, and in each column; only one of the three figures is black in each case. These relations enable us to choose no. 4 as the correct item.

The induction of relations and correlates can also be presented in a verbal form, as in the following test item: black is to white as high is to: green, poor, low, big. Or the presentation can be in a numerical form: 2 is to 4 as 5 is to: 5, 10, 20, 50. Or items can be presented in serial form, as below:

A C F J O __; this an alphabet series; or in numerical form:

2 3 5 8 12 __; this a number series test item. Of the making of such items, there is of course no end; hundreds of thousands have been written and are available in standardized form.

How is acquired knowledge related to intelligence, as so measured? We can write an item measuring knowledge in the same form as we can write an item measuring intelligence, as is shown in the following example: Odysseus is to Penelope as Menelaus is to: Circe – Helen – Nausicaa – Artemis – Eos. Here the relationships involved are too obvious to call for much intelligence, but what is required is a knowledge of whom Menelaus was married to. This can only be regarded as a measure of intelligence when all the people to whom the test is given come from a fairly uniform environment, have gone to similar schools, and been exposed to similar types of teaching. For scientific purposes such an item would be almost useless, except for one fact which is very important, namely that people who are found to be intelligent in terms of the culture-fair type of test we have been discussing so far also tend to acquire knowledge much more readily than persons not so intelligent, and that consequently they will do well also on tests of pure knowledge, such as a vocabulary test measuring one's knowledge of unusual words. What is measured by culture-fair type tests is sometimes called "fluid ability", and what is measured by vocabulary and other types of tests is sometimes called "crystallized ability"; what is important to note is that the two are fairly highly correlated, although there are differences between them.

We have looked at the kinds of item that might go into an intelligence test; how is the test itself constructed? It is sometimes objected by critics that this construction is subjective and is biased by the fact that test authors are usually white and middle-class. Nothing could be further from the truth. As already pointed out, when large numbers of extremely varied test items are intercorrelated on a random sample of the population, all the correlations are found to be positive; in other words, all these test items tend to measure the same thing, but to varying degree. Clearly those items that show the highest correlations are better measures of whatever the total set of items attempts to measure, than would be items having lower intercorrelations. This gives us an objective means of selecting test items; there are of course other, equally objective criteria. Thus *difficulty level* is one; we must have easy, average and difficult items in order to test people of different degrees of intelligence, and statistical considerations tell us how item difficulty should be distributed. Another criterion is *variety*; we should try to introduce as many different types of test item as possible in order to avoid accidental intrusion of irrelevant factors. Different tests, constructed along these principles, correlate quite highly together, giving us confidence that they measure the same thing. Furthermore each of these tests has a high *test-retest reliability*; when applied twice to the same population (or when alternative versions of the test are applied to the same population), correlations of 0.95 are quite usual. When a child or an adult is tested with the Wechsler and then with the Binet Tests (two of the best known individual tests), they do not necessarily get an identical IQ, just as we noted above that different thermometers give slightly different readings; however, the intelligence quotients recorded by these two tests will be almost identical; sufficiently so for most practical and scientific purposes. Greater accuracy still can be achieved by using larger numbers of tests and test items; the relation between accuracy and number of test items is perfectly well understood statistically.

In addition to general intelligence, there are also various group factors – verbal ability, numerical ability, visuo-spatial ability, rote memory, etc. These are much less important, but from the practical point of view they can be very helpful in advising pupils or students as to the best subject to study, or in vocational guidance and occupational selection. Again it is possible to construct specific tests for the measurement of these special abilities, and many such tests are in existence. It is interesting to note that while there are no sex differences in general intelligence, the sexes do seem to differ with respect to some of these special abilities. Thus women tend to be superior with respect to verbal ability, men with respect to visuo-spatial ability. There are good reasons for these observed differences in evolutionary history, as discussed in some detail in [1].

Causes of Individual Differences in Intelligence

There has been much argument about the causes of individual differences in intelligence. Sir Francis Galton believed that these were largely genetic, and there has been much work done to support or disprove this hypothesis. The results are fairly clear-cut, demonstrating that Sir Francis was essentially right in his surmise. I will just mention some of the major methods used to elucidate this problem; modern methods of biometrical genetical analysis are very complex and highly technical, and cannot be explicated here. They are essentially quantitative, depend on analysis of variance designs, and the different methods of experimentation enable us to obtain different estimates of heritability which, if they are to be trusted, should give reasonably similar results. This they do, as I have shown [1], and the general conclusion to be drawn is that heritability accounts for something like 80% of the total variance, environmental factors only 20%. Heritability here includes not only the usual additive genetic gene effects, but also dominance and assortative mating, both of which have been found to be relevant to the inheritance of intelligence. High intelligence is dominant over low intelligence, as we might have expected from simple evolutionary considerations, and assortative mating is widespread, i.e. intelligent men tend to marry intelligent women, and vice versa. Non-additive factors of this kind account for something like 10% of the total variance.

The two most widely used methods of analysis rely on the phenomenon of twinning. The first method bids us study monozygotic twins (twins sharing 100% heredity) who have been brought up in separation from each other, i.e. in different environments; any observed differences in IQ would be entirely due to environmental factors. It has been found that on the whole identical twins of this kind are very similar with respect to intelligence, even though they may be brought up in very different environments. The other method uses a comparison between monozygotic and dizygotic twins, on the assumption that if heredity is important in determining differences in intelligence, then identical twins should be more alike than fraternal twins. This has always been found to be so. Both methods, subject to certain assumptions, enable us to derive quantitative estimates of heritability, and these estimates come out rather similar, and give the results already mentioned.

A third method, also widely used, is to look at adopted children, and see whether in intelligence they resemble their natural parents or their adoptive parents. The outcome of many such studies has been very clearcut, supporting the hereditary view; adopted children are much more like their natural parents than their adoptive parents. Another very powerful argument comes from the so-called regression phenomenon. As Galton already discovered, physical and mental traits which are determined genetically, but less than 100% so, show regression to the mean from the parents to the children. In other words, very tall parents have tall children, but the children are not quite as tall as the parents. Similarly, if the parents are very short, their children will be below average in height, but somewhat taller than their parents. Equally with respect to intelligence – the children of highly intelligent parents will be intelligent but not quite as bright as their parents, and the children of very dull parents will be dull but brighter than their parents – on the average! Figure 2a shows the erroneous picture many people have of the implications of a strong determination by genetic factors of intellectual differences. Here the parents are shown at the top, children at the bottom, and four very dull parents have four very dull children, four very bright parents have four very bright children, sixteen dull parents have sixteen dull children, sixteen bright parents, sixteen bright children, and twenty-four average people have twenty-four average children. The true position is shown in Fig. 2b; regression ensures that there is a considerable mix-up between the two generations. Of the four very dull parents, only one very dull child issues. Two are dull and one is average. Similarly, of the four very bright parents, the issue is distributed as follows: one very bright, two bright, and one average. Of the twenty-four average parents, ten children are average, six are bright, six are dull, one is very bright and one is very dull.

It will be seen that if different social strata are to maintain their mean IQs, then it is essential that there should be a considerable degree of social mobility, with the bright children moving up from the lower classes, and the dull children sinking down from the upper and middle classes. This has indeed been found in all Western cultures, sufficient to counteract the effects of regression. The same effect can be found in single families, with bright siblings rising in the world, dull siblings sinking down into lower socioeconomic status groups. Intelligence has a very important bearing on a person's social standing!

Some critics have objected that regression is merely a statistical phenomenon, and cannot be cited as proof for the importance of genetic factors. This criticism is incorrect. On a purely environmentalist basis we would expect that intelligent parents of high socioeconomic status should give their children the best possible environment, mostly better than the parents themselves enjoy, and that consequently the children should have if anything higher IQs than the parents. Conversely, very dull parents give their children the

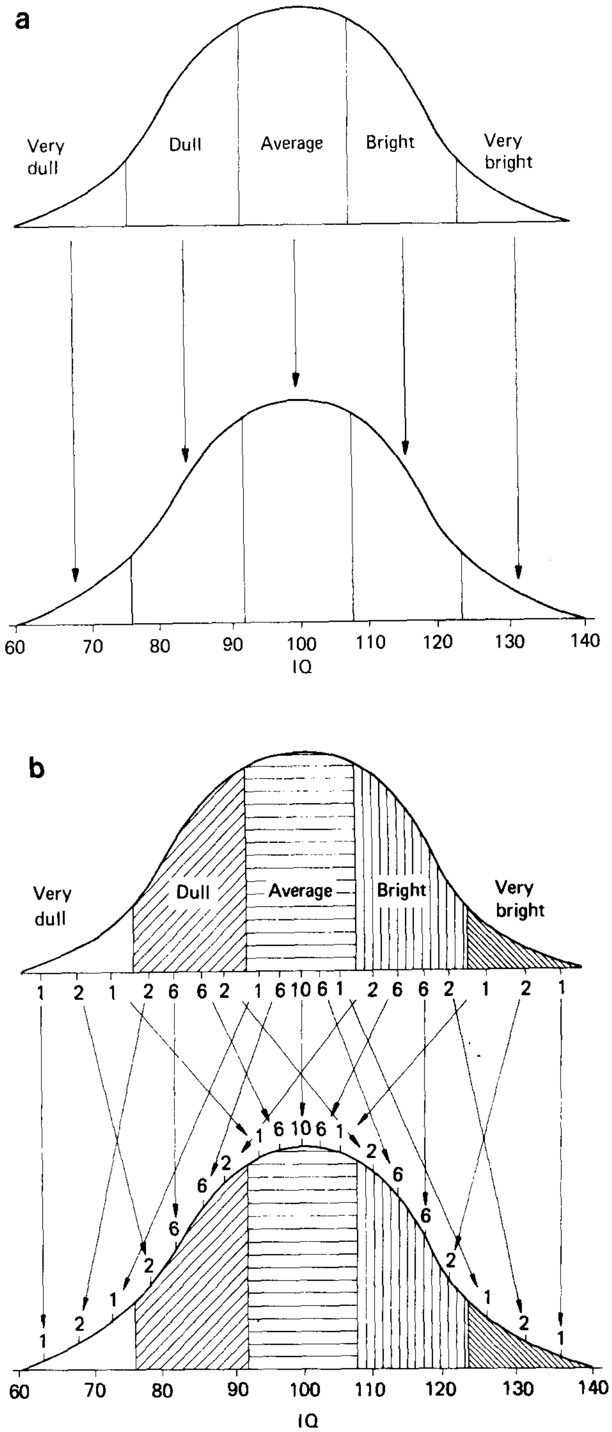

Fig. 2. Illustration of false (a) and true (b) meaning of heritability of intelligence, illustrating regression effects in (P)

very worst possible type of environment, and the children should therefore be very dull, possibly even duller than the parents. The fact that the opposite is found shows that the effect is more than a statistical artefact. Furthermore, there is in existence a genetic formula which enables us to predict the amount of regression to be expected in subsequent generations, given the known degree of additive heritability. If we now substitute in this formula the figure for heritability derived from twin studies, and look at published figures of regression, we find that these are predicted with a high degree of accuracy. It is this type of interaction between different methods of proof, along quantitative lines, that so strongly supports the genetic hypothesis.

These are some of the direct genetic experiments done to demonstrate and measure the degree of genetic determination; there are also more indirect methods used to assess the importance of environmental causes. Thus for instance a recent large-scale experiment was carried out in Poland, where the city of Warsaw had been destroyed almost completely at the end of the Second World War, and had been rebuilt by a Communist government in an attempt to introduce egalitarian principles. People were allocated to identical-type houses, shopped in identical shops, sent their children to identical schools, had identical health services at their disposal, and were treated as alike as possible by a government dedicated to the principle of equality. When the children growing up in this particular environment had their IQs tested by means of a culture-fair test widely used in the West, it was found that the intellectual differences between children so raised in an egalitarian environment were almost as large as those found in children raised in the non-egalitarian environment of the capitalist West; in other words, however egalitarian the environment, children still grow up with very different degrees of intelligence shown on test. Other studies of a similar kind have shown that when environmental influences are studied directly, the quantitative results demonstrate the correctness of the estimate of 20% of the total variance being contributed by environmental factors.

Electrophysiological Measurement of Intelligence

If genetic factors are so prominent in determining differences in IQ between people, then it should be possible to find a biological basis for the observed IQ differences. Our own work, following upon some early experiments in Canada by J.P. Ertl, has made use of the so-called evoked potential on the EEG. Figure 3 illustrates this phenomenon, which is produced by exposing the subject to a sudden stimulus,

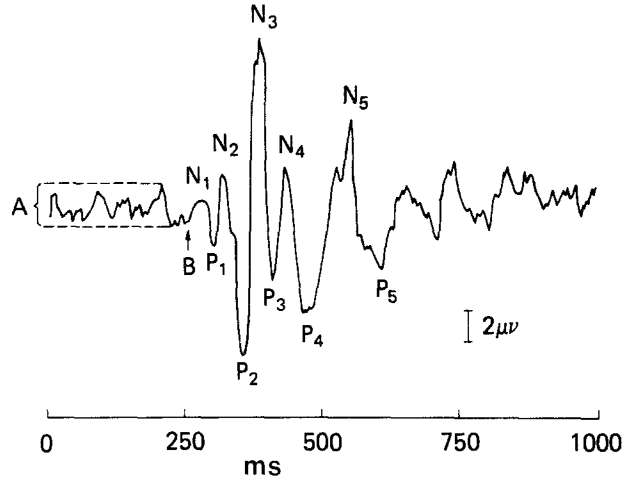

Fig. 3. Illustration of the averaged evoked potential. *A* baseline band, *B* stimulus onset, *N* negative AER component, *P* positive AER component

whether auditory or visual, i.e. a sudden noise delivered over earphones, or a sudden flash of light. In the Figure, *A* shows the base line EEG activity; *B* shows the onset of the stimulus. It will be seen to be followed by a series of waves where successively *N* components indicate negativity, successively numbered *P* components positivity. The evoked potential has a poor signal-to-noise ratio and consequently usually a number of time-locked evocations are averaged in order to produce a final trace. In our own work we have averaged 90 evoked potentials.

Figure 4 shows results for a bright, an average, and a dull subject; in each case the IQ obtained from administration of the Otis Test is included. These waves come from the work of Ertl, and it will be seen that the latency of the waves is greater for the dull subject than for the bright. Intelligent subjects have a quick succession of waves, and the duller they are, the slower do these wave successions become. What would correspond to P5 in these waves occurs roughly after 130 ms for the brighest subject, 350 for the average subject, and 530 ms for the dullest subject. Unfortunately correlations with orthodox intelligence tests are not as high as might be desired, amounting only to about 0.3 at best, as long as we restrict ourselves to visual stimuli and measured latencies.

In our own work we first succeeded in obtaining higher correlations by using auditory stimuli (which are not so subject to artefacts in the EEG measure as are visual stimuli), and by additionally measuring amplitudes – we found that higher IQ subjects had greater amplitudes of the evoked potential waves. By these means we obtained correlations up to 0.6, but even that was unsatisfactory because there was lacking a theoretical underpinning for these essentially

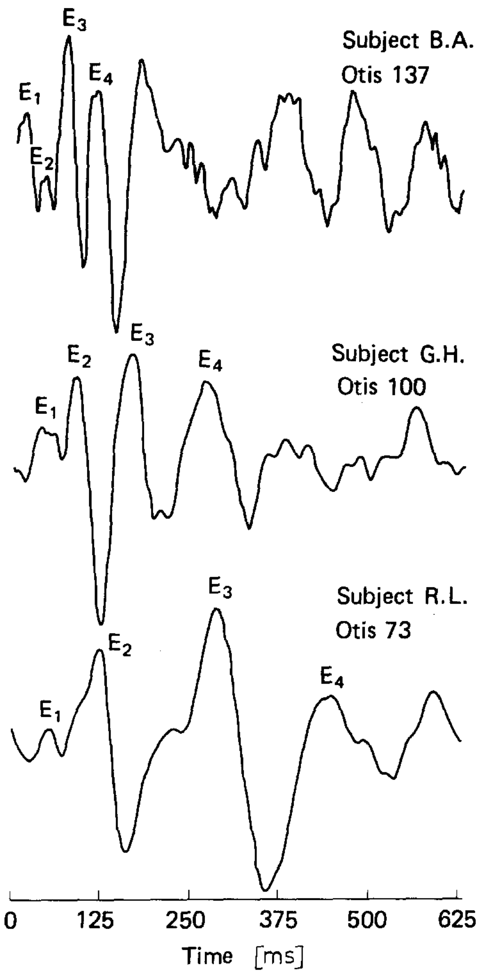

Fig. 4. Averaged evoked potential of three subjects of different levels of intelligence, as indicated by their Otis IQs

ad hoc findings. Recently we have developed a theoretical basis for the measurement of evoked potentials, conceiving of these as being outward records of information processing through neurones and synapses. It is further hypothesized that during this processing of information errors in transmission occur, probably at the synapse. It is further proposed that intelligence is negatively correlated with the probability of errors occurring; in other words, the brighter the person, the fewer errors are likely to occur in transmission of information through his cortex.

From these assumptions (which are elaborated in considerable biochemical and physiological detail in [2]), it may be deduced that the evoked potential trace for intelligent people should be more complex than the trace for dull people, because, as will be remembered, the evoked potential trace is the average of 90 time-locked evocations, and such averaging will preserve smaller details only when these details are repeated properly time and time again. Now when many errors occur, then each individual trace will be different from the others, and summation and averaging will only give us the major outlines of similarities between traces, but not the finer detail.

This hypothesis was strongly supported by a recent and as yet unpublished study by E. Hendrickson, in which she tested 250 school children with the Wechsler Intelligence Test, a well known individual test containing 11 sub-tests; she also obtained averaged evoked potentials from each child, both for the auditory and for the visual stimulus paradigm. Using a measure of complexity of the evoked potential she found a correlation of 0.84 between Wechsler IQ and evoked potential for the auditory stimulus; for the visual stimulus the correlation was somewhat lower, namely 0.73. These correlations are very encouraging indeed; the evoked potential correlates more highly with the Wechsler IQ than does the Wechsler IQ with, say, the Binet IQ, or any other well-standardized and widely used intelligence test. Further statistical calculation showed that when the 11 sub-tests of the Wechsler were intercorrelated with each other and with the evoked potential, the pattern of intercorrelations showed very clearly that the evoked potential measured whatever the Wechsler sub-tests measured, but more reliably, more validly, and more accurately. Figure 5 shows typical evoked potential traces for 6 bright and 6 dull children, together with their Wechsler IQs; readers can see for themselves the very marked differences in complexity of the trace.

It is well known that EEG patterns are very strongly determined by genetic factors, and are difficult to influence by environmental forces; they certainly would seem to be immune to cultural, social and educational factors which have been suggested by some to be responsible for individual differences in intelligence. Clearly the results of these psychophysiological studies strongly support the genetic interpretation of IQ differences, and locate the causal factors

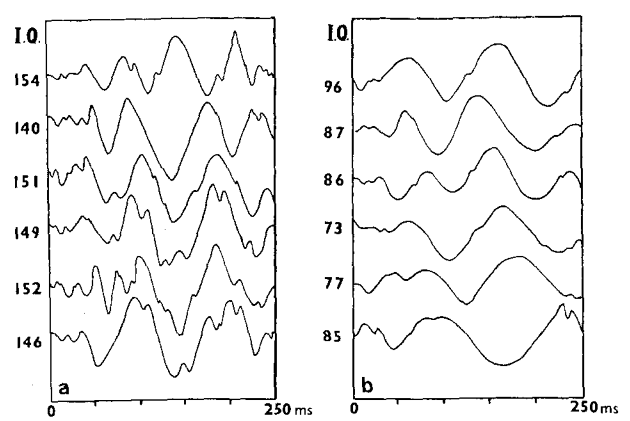

Fig. 5. Averaged evoked potentials of six bright (a) and six dull (b) subjects

firmly in the structure of the human cortex, thus emphasizing their biological determination.

Political and Social Questions

It is unfortunate that the discussion of the environmental and biological factors determining intellectual differences has often been vitiated by political considerations about the desirability or otherwise of egalitarianism; such considerations have no place in science. In any case, there are no grounds for asserting, as some critics have done, that the findings reported here are in any way anti-Marxist. Marx himself was no egalitarian, and fully appreciated, as did Engels and Lenin also, the importance of evolution in accounting for biological differences in ability between persons. Thus modern teaching in the USSR, the DDR and other Communist countries is increasingly getting closer to Western teaching along the lines of this articles. Thus it would be quite wrong to suggest that in some way the debate is between Right and Left Wing, politically motivated groups; questions of politics like this are completely irrelevant to the problem, and whatever the answer to the problem itself may be, it has no bearing on political choices of this type.

One point in particular has often been suggested to discredit the measurement of IQ, namely the empirical finding of large differences between different racial groups. Thus the Chinese and Japanese have often been found to have very significantly higher IQs than Europeans or Americans of Caucasian extraction. It is difficult to account for these differences on socioeconomic or educational grounds, because when the groups measured on IQ are compared with respect to these other factors, the yellow races tend to be of lower socioeconomic status, and to have poorer schooling. Furthermore, the tests on which they prove themselves superior are of course tests constructed by white, middle-class psychologists! Even culturally very deprived groups, such as Eskimoes living in their natural habitat, have been found to be equal on culture-fair tests to white Canadians and Americans having far better education, and living under very different conditions. Such findings must make one suspicious of those who would explain all differences in IQ on purely environmental grounds.

Nothing said in this article should of course be interpreted as suggesting that the last word on these debates has been said, or that there are not still many difficulties, problems and anomalies remaining. Science never reaches an absolute truth; the more we learn about a given subject, the more we realise the extent of our ignorance. However, it is also possible to under-rate what has already been achieved, and it is noteworthy that most of the criticisms of IQ testing, and genetic studies carried out in this field, has appeared in popular rather than scientific journals, and has been made by journalists and politicians, rather than by behavioural geneticists and psychometric psychologists. Very little in the way of proper scientific critique of the stand adopted here has appeared in the scientific literature, and of course this is the only kind of criticism that is scientifically acceptable. There is still debate about the precise degree of heritability of intelligence, and while the figure of 80% here given seems a reasonable one, there is of course a certain margin of error around it which suggests that it will probably not be lower than 70% and not higher than 85%, for the type of population studied. No doubt future work will refine the estimates we have at the moment, but it seems unlikely that such research will alter the findings in any more drastic manner.

1. Eysenck, H.J.: The Structure and Measurement of Intelligence (with D. Fulker). Berlin-Heidelberg-New York: Springer 1979
2. Hendrickson, D.E., Hendrickson, A.E., in: Personality and Individual Differences, Vol. 1, p. 3 (1980)

Received November 10, 1980