# INTELLIGENCE ASSESSMENT : A THEORETICAL AND EXPERIMENTAL APPROACH\*

## By H. J. EYSENCK,

(Institute of Psychiatry, University of London)

### I.—DEVELOPMENT OF A CONCEPT.

ATTEMPTS to measure intelligence have passed through several stages since Galton tried to use the measurement of sensory processes to arrive at an estimate of the subject's intellectual level (1883), and McKeen Cattell 1890) employed tests of muscular strength, speed of movement, sensitivity to pain, reaction time and the like for a similar purpose. These largely abortive efforts were followed by the first stage of intelligence measurement properly so called ; it may with truth be labelled the 'g' phase because both Spearman (1904) and Binet and Simon (1905) stressed the importance of a *general factor of intellectual ability*, Binet contributing mainly by the construction of test items and the invention of the concept of mental age, Spearman contributing mainly by the application of correlational methods and the invention of factor analysis.

The second stage was concerned with the proper definition of intelligence, and theories regarding its nature. Several books concerned themselves with this problem (Thurstone, 1926; Spearman, 1923), and a number of symposia were held (Brit. J. Psychol., 1910; J. Educ. Psychol., 1921; Internat. Congress of Psychol., 1923). Among the theories canvassed were 'mental speed 'hypotheses which placed the burden of intellectual attainment on speed of mental functioning, and 'learning' hypotheses which protested that the ability to learn new material was fundamental. Both hypotheses faced difficulties; the fact that reaction times showed no relation to ability tended to discourage believers in the 'speed' hypothesis, and the negative results of the large-scale work of Woodrow (1946) on the relation between different learning tasks and intelligence discouraged believers in the 'learning' hypothesis. Psychologists learned to agree to disagree, and to present their work with the dictum that "intelligence is what intelligence tests measure "-a saying less circular than it sounds, but only acceptable if all intelligence tests did, in fact, measure the same thing, which they guite emphatically did not.

We thus reach the third stage, which is essentially a continuation of the early factor analytic approach, but now fortified by recourse to multiple factors and matrix algebra. This phase owes most to Thurstone, but Thomson, Burt, Holzinger, and many others made valiant contributions. In this factorial phase, investigators went back to Binet's idea of different mental faculties making up the complex concept of intelligence, and used factor analysis to sort out these alleged faculties ; they emerged with verbal, numerical, perceptual, memory, visuo-spatial and many other factors. At first, Thurstone and his followers believed that these ' primary factors ' put paid altogether to the notion of intelligence, but when they found the primary factors to be themselves correlated they resurrected the concept of intelligence as a second-order factor, a solution already implicit in the earlier methods and theories of Burt (Eysenck, 1939).

\* This paper was originally delivered at a symposium on New Aspects of Intelligence Assessment at the Swansea Meeting of the B.P.S., on 3rd April, 1966. The preparation was assisted by a grant from the M.R.C.

The fourth stage constitutes essentially an extension of the third, and is associated specifically with J. P. Guilford (1966), whose publication of his "1965 model of intelligence" provided some of the motivation for this paper. This model, which shows some similarities to one I published in Uses and Abuses of Psychology (1953, p. 38), is illustrated in Fig. 1. Guilford classifies the intellect into operations which it can perform, different contents of these operations, and different products; by taking all possible interactions we obtain 120 cells corresponding to different mental abilities. Of these Guilford claims to have evidence in actual factorial studies for eighty; he is optimistic about discovering the remainder. To some critics, this factorial extension of Thurstone's work has appeared almost as a reductio ad absurdum of the whole approach. There is a possibility of infinite sub-division inherent in the statistical method employed, and evidence is lacking that further and further sub-factors add anything either to the experimental analysis of intellectual functioning or the practical aim of forecasting success and failure in intellectual pursuits (Vernon, 1965). Worse, the model fails to reproduce the essentially hierachical nature of the data; the one outstanding fact which recurs again and again in all analyses is the universality of positive correlations among all relevant tests, and the positive correlations between different factors (McNemar, 1964). By omitting any mention of this central feature of the scene Guilford has truly cut out the Dane from his production of Hamlet. If this is really the best model (1965 style) which psychology can offer of intelligence and intellect, then the time seems to have come to retrace our steps ; something has gone very wrong indeed !



FIGURE 1. Model of the structure of intellect (Guilford, 1966).

## H. J. EYSENCK

#### II.—LIMITATION OF THE FACTOR ANALYSIS APPROACH.

Zangwill has several times suggested that the whole intelligence testing movement is a technological rather than a scientific one, and in essence my own diagnosis is not too different from his. I would suggest that the psychometric approach has become almost completely divorced from both psychological theory and experiment, and that factor analysis, while an extremely useful tool, cannot by itself bear the whole burden which has been placed upon it. It is the purpose of this paper to raise certain questions in this connection rather than to give definitive answers; a few empirical results from some of our work will be presented more in order to illustrate an approach than bccause we believe that these results settle the questions the experiments were designed to investigate.

#### TABLE 1

FIVE-ITEM INTELLIGENCE TEST, ADMINISTERED TO FIVE CHILDREN ALL HAVING A SCORE OF 2.

	1	2	3	4	5	Total Score
Jones Charles Smith Lucy Mary	R W R R R R	R R A A W	N W A N R	N R R N W	N N A N W	2 2 2 2 2

R—Right answer. W—Wrong answer. A—Abandoned item. N—Item not attempted. (In most tests A and N cannot be distinguished.)

Our work started out with a fundamental criticism of the whole testing movement, directed at the unit of analysis chosen. Nearly all factor analysis and psychometrists correlate test scores and then proceed to work with these correlations; they thus assume that equal scores are equivalent. Such an assumption is unwarranted in the absence of proof, and consideration of typical intelligence test papers shows that it is, in fact, mistaken. Consider Table 1, which shows the results of giving an imaginary five-item test to five candidates. Let R stand for an item correctly solved, W for an item incorrectly solved, A for an item abandoned, and N for an item not attempted. Let us also assume that the items increase in difficulty. It will be seen that all five children obtain an identical mark of 2; but it will also be seen that no two children obtain this mark in the same way. Jones gets the easiest two right, but uses up all his time and does not attempt any more ; he works slowly and carefully. Charles gets some easy items wrong and some difficult ones right; he works quickly but Smith gives up on three items ; had he been more persistent, he carelessly. might have solved some of them. Lucy is rather selective in the choice of item to be tackled, and Mary fails to check her answers, getting three of them wrong. Can it really be maintained that the mental processes and abilities of these five children are identical, merely because they all obtained the same final mark? This is the implicit assumption underlying the factor analysis of test scores, and it may be suggested that this assumption requires careful investigation before we can regard it as acceptable. Such investigations are notable by their absence, and factor analysts proceed throughout as if the problem did not exist. This, it may be suggested, is not a proper scientific procedure.

## III.—THE FURNEAUX MODEL.

Our own approach has been to emphasize the point that the fundamental unit of analysis must be the individual test item, and that in addition to determining the category (R, W, A, N) into which it falls for each candidate, it is important to determine the speed with which each R item is solved, the length of time devoted to each A item (persistence or continuance), and the number of W items together with the time spent on each. Furneaux (1960) has given a detailed analysis of scores obtained in this fashion, and has suggested on the basis of this evidence that the solution of mental test problems has three main parameters: (1) mental speed, i.e., speed of solution of R items; (2) Continuance, or persistence in efforts to solve problems the solution to which is not immediately apparent; and (3) Error Checking Mechanism, i.e., a mental set predisposing the individual to check his solution against the problem instead of writing it down immediately. Two interesting and important consequences follow from this analysis. In the first instance, Furneaux reinstates the mental speed factor to its theoretical pre-eminence as the main cognitive determinant of mental test solving ability, and in the second instance he emphasizes the importance of non-cognitive (personality) factors in determining mental test performance—both persistence and carefulness in checking are personality attributes rather than cognitive abilites. I have attempted to incorporate some elements of this analysis into my own model of intellect (Eysenck, 1953), which is shown in Fig. 2, and which may be compared with Guilford's. What I call mental processes ' he calls ' operations '; what I call ' test material ' he calls ' contents'; so far there is close agreement. But instead of having a third dimension concerned with ' products' (which seems to me a weak and not very important principle of division) I have suggested a dimension rather vaguely labelled 'quality' into which I wanted to incorporate concepts of mental speed and power, somewhat after the fashion of Thorndike's fundamental contribution (1926). The suggestion is that mental speed and power are fundamental aspects of all mental work, but that they are to some extent qualified by the mental processes involved and the materials used. This seems to me a more realistic concept than Guilford's, as well as having the advantage of retaining the central  $g^{\dagger}$  concept in a hierarchical structure in which the



FIGURE 2. Model of the structure of intellect (Eysenck, 1953).

major source of variation is mental speed, averaged over all processes and materials. 'Primary mental abilities,' so called, would then emerge at a lower level of generality, and be related to different processes and different materials used.

Furneaux has demonstrated the fundamental nature of the mental speed function by showing that when an individual's R latencies are plotted against the difficulty level of the items concerned, a negatively accelerated curve is obtained (Fig. 3, A); when the time units are then logarithmically transformed *all plots become linear and parallel* (Fig. 3, B). This may be interpreted to mean that the only source of difference in intellectual ability between individuals (in relation to the particular set of test items chosen at least) is the intercept on the abscissa. The increase in log. latency with increase in item difficulty turns out to have the same slope for all individuals tested, and is thus a *constant*, one of the few which exist in psychology. It seems to me that the scientific study of intelligence would gain much by following up the important leads given by Furneaux in this extremely original and path-breaking work.



FIGURE 3.

Relation between difficulty level of test items and time (A) and log time (B) needed for solution. Alpha, beta and gamma are three imaginary subjects of high, medium and low mental ability, respectively (Eysenck, 1953).

#### IV.-MENTAL SPEED AND INTELLIGENCE.

On the theoretical side Furneaux has suggested that what may be involved in problem solving activity may be some kind of scanning mechanism the speed of which determines the probability of the right solution being brought into focus more or less quickly. If we join this notion with that of information processing, we may have here not only the suggestion of a useful theory of intellectual functioning, but also an argument against those who abandoned the whole theory of 'speed' as underlying intelligence because of the failure of reaction time experiments to correlate with intelligence tests. Let us consider the amount of information conveyed by flashing a light and requiring the subject to press a button located underneath the light flashed. When there is only one light/button combination, no information is, in fact, conveyed. As the number of combinations increases, the amount of information conveyed increases logarithmically, so that one bit of information is conveyed with two combinations, two bits with four combinations and 3 bits with eight combinations. Response speed has been shown by Hick (1952), Hyman (1953) and Schmidtke (1961), to increase linearly with increasing number of bits of information, as shown in Fig. 4 (Frank, 1963). We have two separate items of information for each subject : one is the raw reaction time, as shown by the intercept on the ordinate, the other is the slope of the regression line, i.e., the rate of increase in reaction time with increasing amount of information processed. If intelligence is conceived of as speed of information processing, then simple reaction time, involving 0 bits of information, should not correlate with intelligence, but the slope of the regression line, showing increase of reaction time with amount of information processed, should correlate (negatively) with intelligence: in other words, intelligent subjects would show less increase in reaction time with increase in number of light/button combinations than would dull ones (This is a slightly more precise way of phrasing Spearman's first noegeneti law.) Experimentally, the prediction has been tested by Roth (1964) who demonstrated that while as expected simple reaction time was independent of I.Q., speed of information processing (slope) correlated significantly with I.Q., in the predicted direction. Reaction time experiments, properly interpreted, do not appear to contradict a theory of intellectual functioning based on the motion of mental speed.



FIGURE 4.

Relation between reaction time in seconds and complexity of task, in bits. Data from Merker (1885) and Hyman (1953). (The Hyman data show results before and after practice.) After Frank (1963).

#### V.—LEARNING AND INTELLIGENCE.

The theory that *learning* is basic to intellectual functioning is not necessarily antagonistic to a theory stressing speed; within the more general speed theory we might expect that speedy learning would be characteristic of the bright, slow learning of the dull. In other words, learning would be one of the 'mental processes' sub-divisions in Fig. 2. The early work of Woodrow (1946) was often considered to have disproved such an hypothesis, but his experiments were too simple altogether to throw much light on the problem; it is not adequate to take subjects who are at different stages of mastery and practice on various types of tasks, who are differentially motivated towards thsee tasks, and who vary considerably with respect to the abilities involved in these tasks, and then to correlate speed of learning on these tasks with each other and with I.Q. Improved experimental and statistical methods have given more positive results regarding the relationship between I.Q. and learning (Stake, 1961; Duncanson, 1964).\*

Another argument has often been presented, e.g., by Wechsler; he has pointed out that a learning task such as 'memory span' correlates poorly with the other tests in the W.A.I.S. and does not predict final total score well. Jensen (1964) has argued that this view is based on a neglect of the low reliability of the test as described by Wechsler; this, in turn, can be raised to any height by simply lengthening the (very short) test, or by improving its design, or both. When correlations are corrected for attenuation, Jensen shows that digit span correlates  $\cdot75$  with total I.Q., has a factor loading of  $\cdot8$  on a general factor extracted from the Wechsler tests, is more culture-free than other tests, and can be shown to obey the Spearman-Brown prophecy formula, thus making it possible to increase its relability to any desired degree. The test can be made more predictive of I.Q. by measuring forward and backward span separately, rather than by throwing them together into one score; apparently these two measures are not, in fact, highly correlated and should not be averaged but combined in some multiple correlation formula, if at all.

Jensen has used Digit Span and serial learning experiments of the traditional laboratory kind in an extensive investigation into personality determinants of invididual differences in these tests ; we shall return to this study later. Here it is relevant to mention that he found a multiple correlation of +0.76between learning ability as so measured and college Grade Point Average, a measure of academic standing. When it is considered that this value was obtained in a relatively homogenous group of persons from the point of view of I.Q., and that this correlation is considerably higher than those usually reported with highly regarded I.Q. tests, then it may become apparent why I am suggesting here that we should take seriously the theory relating the concept of 'intelligence' to learning efficiency and speed, and attempt, by means of laboratory studies such as those of Jensen and Roth, to investigate deductions from such an hypothesis. It seems reasonable to expect that such investigations are more likely to help in the elucidation of the nature of intellectual functioning than is the continued construction of I.Q. tests of a kind that has not materially altered in fifty years. And it is also possible that from the practical point of view, this method of procedure may result in tests and devices which enable us to

<sup>\*</sup> An early study showing the close relation obtaining between intelligence, on the one hand, and learning/memory, on the other, was an investigation by Eysenck and Halstead (1945) of fifteen learning/memory tests; these were found to be highly correlated with intelligence. A factorial analysis gave rise to a general factor of intelligence, leaving no residual evidence of any additional contribution by learning or memory.

give better predictions of school and university success than do existing tests.\*

As an example of the much increased possibility of psychological analysis opened by the use of laboratory methods in this field, consider Schonfield's (1965) study of memory changes with age. The general loss of ability of the aged to do I.O. tests well has been known for a long time, as has their failure to acquire new skills and information, or to retain acquired material. These defects may be due either to a loss of ability to retrieve memories from storage, or to a deficiency in the storage system itself. By comparing recall and recognition scores on a learning task, Schonfield showed that recall was impaired in aged subjects, but recognition was not; he concluded that it was retrieval from memory storage which was at fault, rather than storage itself, thus suggesting that learning itself might be unimpaired with age. This experiment is cited, not because the results are definitive in any way, but because they illustrate well the approach suggested here ; simple I.Q. testing cannot in the nature of things do any more than reveal the existence of a deficit, but in order to reveal the precise psychological nature of the intellectual deficit in question more experimental methods are required.

### VI.--LEARNING AND PERSONALITY.

In our discussion of Furneaux's contribution, we found that of his three components of intellectual functioning, only one (speed) was cognitive, while two (persistence and the error-checking mechanism) seemed more orectic in origin, and likely to be related to personality. Most workers in the field of intelligence testing disregard personality factors altogether, but this is almost certainly a mistake. There are several experiments which bring out fairly clearly the importance of personality factors such as neuroticism and extraversion /introversion in the measurement of intelligence, and much of our work has centred on this aspect. Consider first of all the simple learning experiments which we have just discussed; here one can perhaps expect personality to play little if any role. This, however, is not so, and it may be interesting to speculate about the kind of relation which one might expect to find. We may with advantage begin by considering the well-known experiments of Kleinsmith and Kaplan (1963). These authors argued, briefly, that learning is mediated by a consolidation process which takes place after the learned material has been registered, but before it is transferred into permanent memory storage. Consolidation is a function of the state of arousal of the organism; the greater the arousal, the longer and more efficient the consolidation, so that higher arousal leads to better memory in the long run. However, while consolidation is proceeding, it interferes with recall, so that while the consolidation process is going on the highly aroused organism is at a disadvantage. Kleinsmith and Kaplan tested their theory by measuring the amount of arousal (G.S.R. reaction) produced by different paired stimuli; for each subject they then picked the most arousing and the least arousing stimulus pairs and had the subject remember the paired stimulus after presentation of the original stimulus.

<sup>\*</sup> One interesting possibility which is suggested by Jensen's work relates to his finding that serial learning tasks and paired associate learning tasks both correlate with I.Q., but not with each other. In view of the dependence of paired associate learning on verbal mediation, in contrast to the rote character of serial learning, it seems possible to regard serial learning as the prototype of Cattell's 'fluid ' ability, and paired associates learning as the prototype of his ' crystallized ' ability (1964). If this suggestion has any value, it may show the way to the construction of a battery of tests less dependent on cultural factors and training than are most existing I.Q. tests.

## H. J. Eysenck

Recall was arranged at different times after original learning for different groups of subjects, and Fig. 5 shows the results; it will be seen that as expected high arousal words are poorly remembered immediately after learning, but show very marked reminiscence effects, while low arousal words are well remembered immediately after learning, but fade out quickly. There is little doubt of the reality of this phenomenon, which has since been demonstrated several times.



Differential recall of paired associates as a function of arousal level (Kleinsmith and Kaplan, 1963).

In this experiment stimuli were measured and grouped according to their arousing qualities. It is equally possible to group subjects according to their arousability, and I have argued that introverts are characterised by high arousal, extraverts by poor arousal (Eysenck, 1963, 1967). If this theory is along the right lines, we would expect extraverts to behave in the manner of the low arousal words in Fig. 5, and introverts in the manner of the high arousal words. In other words, for short recall times, extraverts should be superior, while for long recall times introverts should be superior. There are about half-a-dozen experiments in the literature demonstrating the superiority of extraverts over short-term intervals, including the work of Jensen already mentioned; these have been summarized elsewhere (Eysenck, 1967), and all that need be said here is that results are in good agreement with prediction. Some unpublished work on pursuit rotor reminiscence also supports the prediction of better learning for introverts after long rest intervals.

A specially designed experiment by McLaughlin and Eysenck was undertaken to test, in addition to the hypothesis stated above, a further one relating

to the personality dimension of neuroticism, which we may regard as associated with drive (Spence, 1964). Subjects were tested on either an easy list of seven pairs of nonsense syllables, or on a difficult list, difficulty being manipulated through degree of response similarity. It was predicted that in accordance with the Yerkes-Dodson law the optimum drive level for the easy list would be higher than that for the difficult list, and it was further assumed that N subjects (high scorers on the N scale of the E.P.I.) would be characterized by higher drive than S subjects (stable, low scorers on the N scale of the E.P.I.). Extraverts, as already explained, were regarded as low in arousal, introverts as high. There are thus four groups of subjects, which, in order of drive, would be (from low to high): stable extraverts; neurotic extraverts and stable introverts; neurotic introverts. (No prediction could be made about the position of the two intermediate groups relative to each other.) The results of the experiment are shown in Fig. 6; extraverts, as predicted, are significantly superior to introverts, and the optimum performance level of drive is shifted towards the low end as we go from the easy to the difficult list, thus shifting the SE group up and the NE group down. (The figures in the diagram refer to number of errors to criterion.)

If introverts, as hypothesized, are characterized by a more efficient consolidation process, due to their greater cortical arousal, then we should be able to predict that they should be superior to extraverts with respect to acquired knowledge. As an example, we may take vocabulary scores, which are clearly the product of learning, and which usually correlate very highly with other I.Q. tests. Eysenck (1947) has reported personality differences between 250 neurotic male soldiers whose Matrices scores were much superior to their Mill Hill Vocabulary scores, and 290 male soldiers whose scores showed a similar difference in the opposite direction; he also studied 200 and 140 neurotic women soldiers showing similar differences. In both sex groups those subjects whose vocabulary was relatively good showed dysthmic (introverted) symptoms, while those whose vocabulary was relatively poor showed hysteric (extra-verted) symptoms. Farley (unpublished) has carried out a study of forty-seven normal subjects in which he found a substantial positive correlation (r = +0.48)between introversion and vocabulary. This is of course in line with the alleged 'bookish' character of the typical introvert. There was no such correlation between Introversion and Raven's Matrices.

It is possible to go further than this and argue that introverts should do rather better at school and university because of this superiority in consolidation of learned material; there is much evidence to indicate that such a prediction may be along the right lines (Furneaux, 1962; Lynn and Gordon, 1961; Savage, 1962; Bendig, 1958, 1960; Otto, 1965; Otto and Fredricks, 1963; Child, 1964; Ranking, 1963a, 1963b). Not all the results are favourable, but the overall impression is certainly in accordance with expectation. It might be suggested that some form of zone analysis (Eysenck, 1966) which included the N variable as well as the E variable would throw much needed light on these relationships. It should be added that the results do not so much support the hypothesis, as rather fail to disprove it. There are so many alternative hypotheses to account for the finding that not too much should be read into the data.

#### VII.—INTELLIGENCE AND PERSONALITY.

It will be clear from this discussion that personality features such as neuroticism and extraversion-introversion interact with learning in complex though meaningful ways, and that great care has to be taken in the design of



Performance of stable extraverts, neurotic extraverts, stable introverts and neurotic introverts on easy and difficult paired associate learning tasks (R. J. McLaughlin and H. J. Eysenck, unpublished data).

experiment not to fall foul of the complex laws relating performance to personality.\* It might be objected that such relations only obtain when laboratory learning tasks are used, but that they fail to appear when orthodox intelligence tests are employed. This is not so. One of the earliest findings relating to extraversion/introversion was that extraverts opt for speed, introverts for accuracy, when there is the possibility of a choice in the carrying out of an experimental task (Eysenck, 1947), and we would expect this difference to appear in relation to intelligence tests also. Jensen (1964) correlated extraversion scores on the E.P.I. with time spent on the Progressive Matrices test and found a significant correlation of -0.46; in other words, extraverts carried out the task more quickly. They also made more errors, but this trend was not significant. Farley (1966) applied the Nufferno test individually to thirty Ss, divided on the basis of their E.P.I. scores into ten extraverts, ten ambiverts and ten introverts. The mean log speed scores on all problems correctly solved for the groups were respectively :  $\cdot 78$ ,  $\cdot 88$  and  $\cdot 93$ . This monotonic increase in solution time with introversion was fully significant by analysis of variance. Other examples of this relation between speed and extraversion are given elsewhere (Eysenck, 1967); 'there seems little doubt about its reality.

Farley (1966) also discovered a significant relation with neuroticism, but as might have been expected (Payne, 1960) this showed a non-linear trend, subjects with average scores being superior to those with high or low N scores. Lynn and Gordon (1961) have also published a study showing a similar trend; they used the Progressive Matrices test. The rationale underlying the prediction of a curvilinear relationship in this context derives, of course, from the Yerkes-Dodson law; it is believed that the optimum drive level for complex and difficult tasks like those involved in an intelligence test lies below the high level reached by high N subjects, and above that reached by low N subjects. The general drive level of the group tested is, of course, quite critical in this connection, and it must be emphasized that unless this can be specified or measured, predictions will not always be fulfilled. Changes in difficulty level of the items, changes in the importance the result of the test assumes in the eyes of the subjects, and changes in the motivational value of the instructions may all lead to a general shift in the drive level of the subjects which may displace the optimum level in either direction. It would seem useful in tests of this prediction to have separate measures of drive, or of arousal, against which performance could be plotted (Eysenck, 1967) : without such direct measures the subjects' N score may often be difficult to interpret, giving us essentially merely a measure of their *probabability* of responding with antonomic activation to an anxietyproducing situation. If the situation is not perceived as anxiety-producing by the subjects, then differences in N cease to matter. This line of argument has led to a better understanding of the conditions under which N correlates with eyeblink conditioning (Eysenck, 1967), and it may be used to design experiments explicitly aimed at increasing the correlation posited.

This dependence of results on precise control of parameter values can also be illustrated by some recent unpublished experiments undertaken by M. Berger. We have noted that extraverts are faster and make more errors when

<sup>\*</sup> The common belief that incentives and higher or depressed motivation generally do not affect intelligence test performance (Eysenck, 1944) may be mistaken; it is conceivable that here too we find the curvilinear Yerkes-Dodson relation, so that *overall* failure to find significant motivation and incentives may be due to compensating positive and negative effects of increased motivation on different types of subject. Some form of interaction terms should be included in analyses of this type, and by preference this should take the form of zone analysis. (Eysenck, 1966.)

conditions are such that the test is administered without stress on speed; in other words, when no explicit instructions are given emphasizing speed, extraverts opt for speed and neglect accuracy, whill introverts opt for accuracy and go slow. These are response styles well familiar from other types of activity (Eysenck, 1960). What would we expect to happen when stress was placed, explicitly and implicitly, on speed of problem solving activity? Let us return to Fig. 6, in which we postulated that stable extraverts would have low drive level. neurotic introverts high drive level, with the other two groups (stable introverts and neurotic extraverts) intermediate. Given the specific stress on *speed* as the proper index of performance, we would expect the low-drive stable extraverts to have the slowest speed, and the neurotic introverts the highest, with the other two groups intermediate : we might also expect that the neurotic introverts speed shown.

Berger tested twenty-one 13-year-old school children in each of the four personality groups : the groups were equated for age, sex and intelligence, using their 11+ records for this purpose. Fifty problems were presented for solution individually, followed by a rest, and finally by another set of thirty problems. Each problem was shown to the child on a screen, with numbered alternative solution; having selected the correct solution, the child pressed a numbered button, which activated a time switch, thus recording solution latency, and also caused the projector to project the next problem on to the screen. Instructions emphasized speed of working, and the whole experimental set-up added to this impression; furthermore, the disappearance of the problem after the button had been pressed eliminated the possibility of checking the correctness of the answer. Figure 7 shows the results of the first fifty items; the next thirty showed similar results. The Figure is arranged in the form of a cumulative time record, with time arranged along the ordinate and the problems, 1 to 50, along the abscissa. It will be clear that the stable extraverts are much the slowest, the neurotic introverts much the fastest, with the other two groups intermediate; these differences are highly significant. It was also found, at a high level of significance, that neurotic introverts compensated for their speed by making more errors than the other groups. Thus, the Yerkes-Dodson law appears to be working here very much as it did in the case of the McLaughlin-Eysenck experiment : the low drive SE group does poorly because it is so slow, the high drive NI group does poorly because it makes too many errors, and the intermediate NE and SI groups do best because they work at an optimum level of motivation.

#### VIII.—FLUENCY AND INHIBITION.

This study illustrates the value of applying theories and laws from general and experimental psychology to intelligence testing. Another example may serve the same function. From the point of view of the experimental psychologist, a typical intelligence test is a good example of a task undertaken in the condition of massed practice; we would, therefore, expect it to generate reactive inhibition. Extraverts generate such inhibition more strongly and more quickly than do introverts (Eysenck, 1957, 1967), and consequently we would expect that when groups of extraverts and introverts are matched for performance during the earlier part of an intelligence test, then they will diverge towards the latter part, with the introverts superior in performance. Another way of saying the same thing would be to regard an intelligence test as a vigilance test, and use the well-known fact that introverts preserve vigilance better than extraverts



FIGURE 7.

Solution times of stable extraverts, neurotic extraverts, stable introverts and neurotic introverts on fifty intelligence problems, timed separately and cumulated. (M. Berger, unpublished data.)

to predict their better performance towards the end. Eysenck (1959) has reported such an experiment, in which he used sixty items from the Morrisby Compound Series test, individually but unobtrusively timed. Using speed of correct solutions, it was found that on the first forty-five problems introverts were slower than extraverts, but on the last fifteen items, the two groups reversed position and the extraverts were now the slower. On the last fifteen items, it was also found that the extraverts gave up more easily. It would thus seem true to say that extraverts do show the predicted decline in performance during the latter part of their performance on a typical test of intelligence, administered as far as the subjects were concerned in the usual manner, and without any special stress on speed. This experiment, taken in connection with the others already quoted, leaves little doubt that personality plays an important part in intelligence test performance, and that its influence has hitherto been very much under-estimated.

Personality factors interact with intelligence test performance in many ways, and neglect of these factors may easily lead to quite incorrect conclusions. As an example, we may, perhaps, take the large body of work recently done on convergent and divergent types of tests (Hudson, 1966). In studies of this kind, candidates good on divergent tests are often called ' creative,' and the argument is sometimes extended to other desirable qualities of intellect, such as ' originality' (Barron, 1963; Taylor and Barron, 1963). In fact, divergent tests are by no means new; under the title of 'fluency' tests they were among the early discoveries of the London school, and a typical set of such tests is reprinted in Cattell's (1936) Manual of Mental Tests. Tests of this kind were found to be correlated with extraversion (Eysenck, 1960) and Spearman (1926) already pointed out that this particular factor "has proved to be the main ground on which persons become reputed for 'quickness' or for 'originality'." Hudson's work supports some such interpretation quite strongly ; ' divergent ' schoolboys, as compared with ' convergent ' ones, are more fluent, make more errors on orthodox tests, are emotionally more forthcoming, are more sociable, and prefer 'arts ' to ' science ' subjects-all characteristics of extraverts as compared with introverts. There is, in fact (as Hudson acknowledges) no evidence to show that 'divergent' boys are more creative than 'convergent' ones; as he points out, one can be 'creative' in different ways. All that we seem to be dealing with in this distinction would seem to be a kind of response set or ' style'; it is, perhaps, unusual to apply this concept in relation to intelligence tests. but it applies here probably more than in relation to personality inventories.

## IX.—The limitation of psychometry.

These various ways in which personality and intelligence testing interact do not by any means exhaust the available evidence. Factor analysts usually assume, without proof, that groups which do not differ in performance on a group of tests will also not differ in factorial solution. Lienert (1963) showed that this assumption is, in fact, erroneous; children high and low on N, respectively, do not produce identical correlation matrices or factors, when administered sets of intelligence tests, nor do the two groups even agree in the number of factors produced. As Eysenck and White (1964) have shown in a re-analysis of the data, "the stable group has a more clearly marked structure in the cognitive test field than has the labile group." (It has also been found that students differing in intelligence do not have identical factor patterns on personality questionnaires; the evidence is presented by Shure and Rogers, 1963). It is not unlikely that some of the observed differences in factor structure are connected with the intellectual response styles which we have found to be characteristic of different personality groups, but at present there is no evidence to indicate precisely how this may have come about. Much further work is clearly required before we can be sure of our facts in this complex field.

All that has been said in this paper is only suggestive, and I do not in any way believe that the hypotheses stated, and sometimes supported by experimental data, are at the moment anything but guideposts pointing in the direction

of interesting and important factors which will almost certainly have a bearing on the proper measurement of intelligence. We have noted four stages in the development of intelligence tests; it is the main purpose of this paper to suggest the importance of starting out on a fifth stage of intelligence assessment, a new stage based on theoretical and experimental work, and not divorced from the main body of academic psychology. Psychometrics and factor analysis have important contributions to make, but they can do so only in conjunction with other disciplines, not by 'going it alone.' What is required is clearly an integration of intelligence testing with the main stream of academic psychology, and a more determined experimental and laboratory approach to the problems raised by the various theories of intellectual functioning. Some obvious suggestions emerge from the inevitably somewhat rambling and unco-ordinated discussion of this paper. (1) Analysis of performance should always take into account individual items, rather than tests, i.e., averages taken over what may be, and usually are, non-homogeneous sets of items. Such analysis should be made in terms of latencies, i.e., speed of individual item solution, as well as of errors, persistence before abandoning items, and other similar differential indicators of response style. (2) Investigators should pay more attention to laboratory studies of learning and memory functions, of speed of information processing, and other experimental measures in the testing of specific hypotheses regarding the nature of intellectual functioning. Analysis of intelligence tests of the orthodox kind raises problems, but cannot in the nature of things go very (3) Investigators should experiment with variafar towards answering them. tions in experimental parameters, such as rest pauses, time from end of learning to recall, rate of presentation, degree of motivation, etc., in an effort to support or disprove specific theoretical predictions regarding the process of learning and problem solving. (4) Personality variables, such as stability-neuroticism and extraversion-introversion, should always be included in experimental studies of intellectual functioning, because of their proven value in mediating predictions and their interaction potential in all types of learning and performance tasks. Vigorous research along these lines carries with it the promise that notions such as intelligence, I.Q., ability and factor will cease to be regarded as poor relations, and will return to the eminent and successful status they held before the war; it also furnishes the only means of making these concepts scientifically meaningful, academically respectable, and practically more useful.

#### X.—References.

- BARRON, F. (1963). Creativity and Psychological Health. Princeton: Nostrand. BENDIG, A. W. (1958). Extraversion, neuroticism and verbal ability measures. J. Consult. Psychol., 22, 464.
- BENDIG, A. W. (1960). Extraversion, neuroticism and student achievement in introductory psychology. J. Educ. Res. 53 263-267. BINET, A., and SIMON, R. (1905). Méthodes nouvelles pour le diagnostic du nivean
- intellectuel des anormaux. Annee Psychol., 11, 191-244.

CATTELL, J. MCK. (1890). Mental tests and measurements. Mind, 15, 373-380. CATTELL, R. B. (1936). A Guide to Mental Testing. London : University of London Press. CATTELL, R. B. (1964). Fluid and crystallized abilities. In R. B. Cattell, Personality

CATTELL, R. D. (1907). This and one of the second se

performance in school examinations. Brit. J. Educ. Psychol., 34, 187-196. DUNCANSON, J. P. (1964). Intelligence and the Ability to Learn. Princeton, N.J.: Educational Testing Service.

EYSENCK, H. J. (1939). Primary mental abilities. Brit. J. Educ. Psychol., 9, 270-275. EYSENCK, H. J. (1944). The effect of incentives on neurotics, and the variability of neurotics as compared with normals. Brit. J. Med. Psychol., 20, 100-103.

EYSENCK, H. J. (1947). Dimensions of Personality. London : Routledge and Kegan Paul.

EYSENCK, H. J. (1953). Uses and Abuses of Psychology. London: Pelican.

EYSENCK, H. J. (1957). The Dynamics of Anxiety and Hysteria. London: Routledge and Kegan Paul.

EYSENCK, H. J. (1959). Personality and problem solving. Psychol. Rep., 5, 592.

EISENCK, H. J. (1969). Fersonality and problem solving. Psychol. Rep., 5, 592.
EYSENCK, H. J. (1960). The Structure of Human Personality. London: Methuen.
EYSENCK, H. J. (1963). The biological basis of personality. Nature, 199, 1031-1034.
EYSENCK, H. J. (1966). Personality and experimental psychology. Bull. Brit. Psychol.
Soc., 62, 1-28.

EYSENCK, H. J. (1967). The Biological Basis of Personality. New York : C. C. Thomas.

EVSENCK, H. J., and HALSTEAD, H. (1945). The memory function. Amer. J. Psychiatry, 102, 174-180.

EVSENCK, H. J., and WHITE, P. O. (1964). Personality and the measurement of intelligence. Brit. J. Educ. Psychol., 34, 197-202.

FARLEY, F. H. (1966). Individual differences in solution time in error-free problem solving Brit. J. Soc. Clin. Psychol., 5. To appear.

FRANK, H. (1963). Informations psychologie and nachrichtentechnik. In Weiner. N., and Schade, J. P. Progress in Brain Research, Vol. 2, Nerve, Brain and Memory Models. Elsevier Publishing Co., Amsterdam, 79-96. FURNEAUX, W. D. (1960). Intellectual abilities and problem solving behaviour.

In H. J. Eysenck (Ed.) Handbook of Abnormal Psychology. London : Pitman.

The psychologist and the university. FURNEAUX, W. D. (1962). Univ. Quart., 17, 33-47.

GALTON, F. (1883). Inquiries Into Human Faculty and its Development. London : Macmillan.

GUILFORD, J. P. (1966). Intelligence: 1965 model. Amer. Psychol., 21, 20-26. HICK, W. (1952). On the rate of gain of information. Quart. J. Exp. Psychol., 4, 11-26. ниск, w. (1952). On the rate of gain of information. Quart. J. Exp. Psychol., 4, 11-26. HUDSON, L. (1966). Contrary Imaginations. London: Methuen. Нуман, R. (1953). Stimulus information as a determinant of recation time. J. Exp. Psychol., 45, 188-196.

JENSEN, A. R. (1964). Individual Differences in Learning: Interference Factors. U.S. Depot. of Health, Education and Welfare. Co-op. Project No. 1867.

KLEINSMITH, L. J., and KAPLAN, S. (1963). Paired-associate learning as a function and interpolated interval. J. Exp. Psychol., 65, 190-193.

LIENERT, A. A. (1963). Die Faktorenstrukter der intelligenz als Funktion des Neurot-izismus. Zhstr. F. Exp. Angew. Psychol., 10, 140-159. LYNN, R., and GORDON, I. E. (1961). The relation of neuroticism and extraversion to

intelligence and educational attainment. Brit. J. Educ. Psychol., 31, 194-203.

MCNEMAR, R. (1964). Lost : our intelligence ?: Why? Amer. Psychol., 19, 871-882. MERKEL, J. (1885). Die zeitlichen Verhältnisse der Willenstätigkeit. Philos, Studien, 2, 73-127. Otro, W. (1965). Inhibition potential in good and poor achievers. J. Educ. Psychol.,

56, 200-207.

OTTO, W., and FREDRICKS, R. C. (1963). Relationship of reactive inhibition to reading skill achievement. Educ. Psychol., 54, 227, 230.

PAYNE, R. W. (1960). Cognitive abnormalities. In H. J. Eysenck (Ed.) Handbook of Abnormal Psychology. London: Pitman. RANKIN, G. F. (1963a). Reading Test Performance of Introverts and Extraverts. Milwaukee

12th Yearbook of Nat. Reading. Conf.

RANKIN, G. F. (1963). Reading test reliability and validity as function of introversion: extraversion. J. Devel. Reading, 6, 106-117.

EYSENCK, H. J. (1967). The Biological Basis of Personality.

ROTH, E. Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenharg mit intelligenz. Atschr F. Exp. Angew. Psychol., 11, 616-622.

SAVAGE, R. D. (1962). Personality factors and academic performance. Brit. J. Educ. Pschol., 32, 251-252.

SCHMIDTKE, H. (1961). Zur Frage der informationstheoretischen Analyse von Wahlreak-tionsexperimenten. Psychol., Forschung, 26, 157-178.
 SCHONFIELD, D. (1965). Memory changes with age. Nature, 208, 918.

SHURE, G. H., and ROGERS, M. S. (1963). Personality factor stability for three ability levels. J. Psychol., 55, 445-456.

SPEARMAN, C. (1904). "General intelligence" objectivity determined and measured. Amer. J. Psychol., 15, 201-293.
 SPEARMAN, C. (1923). The Nature of "Intelligence" and the Principles of Cognition. London: Macmillan.

SPEARMAN, C. (1927). The Abilities of Man. London: Methuen.

SPENCE, K. W. (1964). Anxiety (drive) level and performance in eyelid conditioning. PSYCHOL., BULL., 61, 129-139.
 STAKE, R. E. (1961). Learning parameters, aptitudes and achievements. Psychometric

Mon., No. 9. TAYLOR, C. W., and BARROW, F. (Eds.) (1963). Scientific Creativity. London : Wiley.

THORNDIKE, E. L. (1926). The Measurement of Intelligence. New York : Teachers' College Columbia University.

THURSTONE, L. L. (1926). The Nature of Intelligence. New York : Harcourt, Brace and World.

VERNON, P. E. (1965). Ability factors and environmental influences. Amer. Psychol., 20, 273-733.

WOODROW, H. (1946). The ability to learn. Psychol. Rev., 53, 147-158.

(Manuscript received 21st July, 1966)