A Model for Intelligence

Edited by H. J. Eysenck

With Contributions by

M. Berger C. R. Brand I. J. Deary
H. J. Eysenck M. K. Gardner A. E. Hendrickson
D. E. Hendrickson A. R. Jensen R. J. Sternberg
P. O. White

With 57 Figures and 51 Tables

Springer-Verlag Berlin Heidelberg New York 1982

Editor
Professor Hans J. Eysenck
Institute of Psychiatry
De Crespigny Park, Denmark Hill
London, SE5 8AF, Great Britain

ISBN-13:978-3-642-68666-5 e-ISBN-13:978-3-642-68664-1
DOI: 10.1007/978-3-642-68664-1

Library of Congress Cataloging in Publication Data
Main entry under title: A Model for intelligence.
Bibliography: p. Includes indexes. 1. Intellect. I. Eysenck, H.J. (Hans Jurgen), 1916–
BF431.M558 1982 153.9 82-5859
ISBN-13:978-3-642-68666-5 (U.S.) AACR2

Softcover reprint of the hardcover 1st edition 1982

2126/3321-543210

Contents

List of Contributors

Dr. M. Berger
Department of Child Psychiatry
St. George's Hospital
Blackshaw Road
London SW17 OQT, Great Britain

Mr. C. Brand
Department of Psychology
University of Edinburgh
7, George Square
Edinburgh EH8 9JZ, Great Britain

Mr. I.J. Deary
Department of Psychology
University of Edinburgh
7, George Square
Edinburgh EH8 9JZ, Great Britain

Professor H.J. Eysenck
Department of Psychology
Institute of Psychiatry
De Crespigny Park, Denmark Hill
London SE5 8AF, Great Britain

Dr. M.K. Gardner
Department of Psychology
Box 11A Yale Station
Yale University
New Haven, Connecticut 06520, USA

Dr. A.E. Hendrickson
15 Lakeside Drive
Esher, Surrey, Great Britain

Dr. D.E. Hendrickson
15 Lakeside Drive
Esher, Surrey, Great Britain

Professor A.R. Jensen
University of California
Institute of Human Learning
Berkeley, California 94720, USA

Professor R.J. Sternberg
Department of Psychology
Box 11A Yale Station
Yale University
New Haven, Connecticut 06520, USA

Dr. P.O. White
Department of Psychology
Institute of Psychiatry
De Crespigny Park, Denmark Hill
London SE5 8AF, Great Britain

1 Introduction

H.J. Eysenck

The concept of 'intelligence' has been with us for a long time. Some two thousand years ago, Plato and Aristotle singled out *cognitive* from *orectic* factors in behaviour, and Cicero coined the term *'intelligentia'*, which has since assumed such universal acceptance. It is only in recent years, of course, that psychologists have attempted to define the concept more closely, to carry out experiments, and to try and measure it. The result has been rather curious. On the one hand we have the overwhelmingly successful application of measures of IQ in education, industrial selection, vocational guidance, officer selection, and many other areas. On the other we have large-scale criticism of concept and measurement alike, including firm denials that intelligence 'exists' at all, or can conceivably be measured. Many of the evils that beset our society, like intellectual differences between races or social classes, are laid at the door of the psychologist, who measures (but can hardly be accused of causing!) these differences. Thus the very success of intelligence testing seems to have caused the storm of criticism that is at present all but submerging it.

In part, this storm is not unrelated to differences in Weltanschauung, Zeitgeist, and Ideology; this is not the place to deal with attitudes and values tangential at best to scientific truth. But it would be absurd to deny that such unscientific aspects of a given theory may play a powerful part in its acceptance or rejection; the history of atomic theory in some ways parallels that of intelligence, as will be pointed out in more detail presently. But before touching upon this aspect, it may be useful to draw attention to another reason for the present unsatisfactory state of the concept of intelligence. Cronbach (1957), in his well-known presidential address to the American Psychological Association, referred to the two disciplines of scientific psychology, meaning the experimental and the correlational. He advocated their unification; only by joining together in the attempt to build a truly scientific psychology could such an objective be achieved.

My own view has always been the same (Eysenck 1967a), and in the companion volume to the present one, I have tried to show how such a unification can be brought about in the attempt to construct *A Model for Personality* (Eysenck 1981a). The book here presented attempts to take the first few, faltering steps in the same direction for intelligence; it is based on the view that hitherto the definition and measurement of intelligence have been the province of the psychometrist, using essentially correlational methods, and that this is not, and can never be enough if we are to gain a proper understanding of this important part of our mind. Correlational methods have to be supplemented by experimental designs; theories suggested by psychometric investigations have to be subjected to univariate analysis; above all, multivariate studies have to take into account theoretical analyses of experimentalists, based on laboratory investigations. It is the absence of such integration which, I am convinced, is responsible for much of the lack of glamour which intelligence testing holds for the experimentalist.

In actual fact, the distinction between experimental and correlational, or between causal and psychometric, is less clear than might appear at first. In psychology, the distinction is obvious enough, and there is seldom any difficulty in assigning a given practitioner to one group or the other. But the principle is less clear when looked at in the light of the philosophy of science. Hume already pointed out the difficulties which the notion of 'cause' encounters when subjected to serious investigation, arguing that essentially all we ever find in our scientific investigations are correlations, not 'causes'. Again, if we agree with Popper that we can never *prove* a theory, but only *disprove* it, then clearly we can never arrive at a causal theory which can be regarded as fundamentally *true*. Perhaps correlation is all we ever observe, all the rest being convenient fiction?

A look at the 'hard' sciences may equally disabuse us of the notion that correlation and causal analysis are clearly separated. Consider Hubble's constant, perhaps the fundamental concept in modern cosmology. V.M. Slipher, at the Lowell Observatory, had found in 1912 that the radial velocities of numerous nebulae showed a clear-cut Doppler effect (red shift), thus suggesting a very rapid motion of recession. Hubble in 1929 showed that these velocities of recession increased proportionally to the distance of the observed nebula. This discovery (essentially a correlation!) implied the general recession of the galaxies, and hence the expansion of the universe. Later work by Hubble and Humason, published in 1953, extended this relation to ever vaster distances, using the Palomar 200″ to extend the reach of the Mt. Wilson 100″ telescope. The original estimate of the Hubble constant was found to be wrong by a whole order of magnitude, but this in no way affected the acceptance and importance of the original discovery. Thus in astronomy there was an intimate relation between a simple correlational finding (stellar distance versus red shift), and theoretical-causal analysis leading to experimental studies of various kinds. Why is there such an opposition between these two methods of working in psychology and why is this opposition apparently missing in the hard sciences?

Fundamentally, the methods of experimental psychology are just as correlational as are those of psychometry. Mathematically, the methods of correlation, partial correlation, and multiple correlation are equivalent to those of analysis of variance and covariance, although different assumptions may be made in particular cases. To say, as do experimentalists, that $a = (f)\ b$, i.e. that the dependent variable, a, is a function of the independent variable, b, is simply to say that they are correlated, with the size of the main effect indicating the size of the correlation. The nature of the experiment may indicate the direction of causal influences, but this can often be done in the case of correlations also. In an experiment we can manipulate the independent variable, but the same can be done in correlational studies, e.g. by varying the range, selection of subjects, or imposing certain types of control. What then is it that so profoundly differentiates the experimental from the psychometric approach, and that so clearly separates psychology from the hard sciences?

The answer, I would suggest, lies in the essential definition of psychology as the science of *behaviour*. Behaviour is only shown by organisms, and the study of the behaviour of organisms by definition makes it impossible to carry out the type of experimental work characteristic of the physicist or the chemist, i.e. the elimination of disturbing influences, and the truly univariate study of a given dependent variable, a, as a function of a given independent variable, b. Experimentation consists essentially in the manipulation of a great number of extraneous variables in order to eliminate their influence on the functional relationship under examination, leaving open the way to the establishment of a clear-cut relationship (a correlation approximating unity) between a and b. Figure 1 shows the results of the Hubble and Humanson study mentioned earlier, with the logarithm of the stellar velocity in thousands of km/s plotted on the ordinate, and the distance, estimated in terms of the photovisual magnitude of the 10th nebula in a given cluster, on the

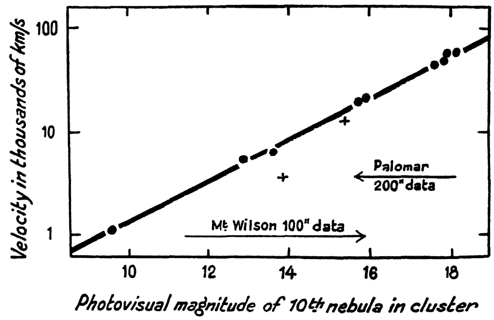


Fig. 1. The velocity-magnitude relation for nebulae

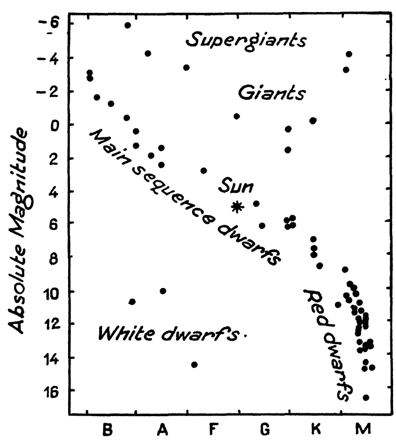


Fig. 2. The Hertzprung-Russell diagram, giving the distribution of stars as a function of their luminosity (absolute magnitude) and temperature (spectral type)

abscissa. The correlation is very close to unity, and hence the relation established by observation is not referred to as a correlation; nevertheless that is of course precisely what it is, with the regression line now the major interest. It is the aim of the physicist to establish correlations of unity, by the elimination of disturbing variables, and this aim is so frequently achieved that the correlational nature of the established relationships is often forgotten.

It is not implied of course that all stars, or nebulae, or all galaxies, are in any sense identical; clearly they are not. The Hertzsprung-Russell diagram (Fig. 2) shows the distribution of stars as a function of their luminosity (absolute magnitude) and temperature (spectral type), outlining their evolutionary development. There is now a typology of galaxies, listing them according to their appearance. But these 'individual differences' aspects of astronomical studies are irrelevant to the observed correlation first described by Hubble; all astral bodies alike obey the Hubble law, and form part of the expanding universe.

Psychology in the nature of things cannot eliminate variables in the same fashion as physics and astronomy because these variables are ineluctably connected with the organism. Different people are characterized by an enormous list of individual differences, ranging from physique through health and strength, to biochemical and physiological differences, and finally to differences in intelligence, abilties, personality, mood, and any other psychological functions, such as sensory thresholds, j.n.d.s., and whatnot that anyone may like to mention. These variables cannot be eliminated, simply because the individual brings them with him into the laboratory; what can of course be done, and is usually done by experimentalists, is to disregard such differences and pretend that all subjects are monozygotic twins, identical for all practical purposes with each other, and with humanity outside the laboratory. This pretence makes it easier for the experimentalist to believe that he is imitating the procedures of physics, but of course such a pretence has to be paid for. The Danegeld consists in the accumulation of variance in the error term, and the small size of the main effects usually observed; experimentalists have sold their birthright for a mess of pottage.

Other consequences which follow from the intrusion of the organism into the most well-planned experiments are the frequency with which regressions are curvilinear rather than linear; the small size of observed correlations (corresponding to the small amount of variance attributable to the main effects); and the difficulty psychologists notoriously have in replicating their observations – any replication involves different organisms, and that alone may change the outcome!

The answer to this problem is of course to take into account the nature of the organism in every experiment involving human beings (or rats, for that matter – the literature is full of experimental demonstrations that different strains of rats may react quite differently to identical experimental situations). In other words, we must take seriously the now quite popular admission that the S – R formula has failed, and must be replaced by another formula incorporating the organism: S – O – R. This leads to a change in the functional prescription as well; we must now write: $A = (f)\ b$, P, where P stands for personality variables of the most variegated kind, including intelligence, physique, etc. This inclusion of personality in typical psychological experiments of course also requires personality theorists to formulate their concepts in terms of variables which have meaning to the experimentalist (Eysenck 1981 a); only in this way can we establish a meaningful relationship between experimentalists and psychometrists.

If experimentalists have been remiss in their failure to establish a working relation with psychometrically inclined psychologists, the latter too have been similarly averse to legitimizing their work by reference to experimental procedures. Where a dialogue could have been so beneficial to both sides, there has been nothing but frosty silence, and occasional ill-considered criticism. Thus the main burden of Cronbach's review of the situation, and advice for improvement, remains unchallenged; many have paid lip service to his wisdom, but few have tried to carry out his recommendations! It is sad to say that work in the field of intelligence has been greatly impoverished by concentrating narrowly on correlational and factor analytic studies (in the most restricted sense); had Thorndike's (1927) early advice been followed, and more theory-oriented experimental studies been done, we might know a great deal more about this mysterious concept than we do now.

To say this is not to disregard the important work done by factor analysts, from Spearman, Burt, and Thurstone to Cattell, Horn, and Guilford. My position has always been intermediate between those who believe that factor analysis is both a necessary and a sufficient methodology for establishing the facts in this area, and those who believe it to be neither necessary nor sufficient! I hold that factor analysis is a necessary but certainly not a sufficient technique for any of the purposes for which it was designed. It furnishes us with important information which it would be difficult or impossible to obtain in any other way; it can be made to test certain hypotheses, although not usually as clearly as we might wish; and it can serve in an appropriate way to establish the number of dimensions (explanatory concepts) required for a given area. But it also suffers from a nearly fatal disease; it does not give rise to agreed solutions, but rather results in a large (in fact unlimited) number of alternative solutions, all of which are mathematically equivalent, and between which there is no objective choice. Factors, however extracted, can be rotated orthogonally or obliquely into an unlimited number of positions; variance can be concentrated on a few (or just one) factor(s), or smeared over a large number. Not in such a manner are truly scientific concepts achieved! (Revenstorff, 1978).

Consider the present position, over 50 years after Spearman (1927) wrote his great book. On one side we have authors like Guilford (1967), who holds that there are at least some 120 independent factors in the ability field, organized into his model of the intellect, and all quite orthogonal to each other, with no general ability (intelligence) to be discerned. On the other hand we have authors like Cattell, Horn, and Vernon, who strongly advocate the recognition of a general factor of intelligence, together with a small number of primaries (Eysenck 1979). It is of course possible to point out that Guilford's position is very difficult to defend, because it relies for its support very much on entirely subjective criteria of rotation ('Procrustes' methods), but mathematically there can be no argument that he would be quite within his rights to distribute the available variance over 120 (or any number) of factors, rather than concentrate it on one single, or a few, factors. Thus it would seem that factor analysis is not capable of deciding issues of fundamental importance, which have been fought over for 50 years or more. Clearly something else is required; we need to formulate theories. Factorial theories are not so testable, unless we impose arbitrarily certain statistical limitations (orthogonal simple structure, for instance) on the solution; but this would be simply begging the question.

To many experimentalists, the very existence of such questions, and the failure of psychometrists to answer them, seems an absurdity. Yet even the hard sciences have encountered similar problems, which, although apparently simple and straightforward, posed great difficulties and led to much argument over prolonged periods. The battle about phlogiston is but one such example, ranging such men as Priestley against others, like the great Lavoisier; the question of the nature of light (wavular or corpuscular) is another. Perhaps nearest to the problem: Does intelligence exist? comes the question: Does the atom exist? Both intelligence and the atom are, of course, concepts, and cannot exist in the simple sense that a table, or a pig, might be said to exist (although even there philosophers might enter a demur concerning the lighthearted use of the verb 'exist'). In both cases, what appears at first a very simple question turned out to be in fact very difficult to answer. And in both cases there seems little doubt that differences in Weltanschauung played an important part in the attitudes adopted towards these two concepts.

Bernal (1969), in his history of science, credits Democritus with the original presentation of an atomic theory, and presents this as 'the most effective answer to (the) idealist tendencies' of men like Parmenides, Zeno, and Plato; this early connection of philosophy with science was to be continued later in the theories of Mach, Ostwald, and others. Democritus, instead of thinking (as had done his predecessors) of a universe of ideal numbers, conceived of it as made up of innumerable small uncuttable (a-tomos) particles, atoms moving in the void of empty space. These atoms were unalterable, of various geometric forms (to explain their capacity for combining), and their movement accounted for all visible change.

Atomic theory was made part of chemistry at the beginning of the nineteenth century, through the work of Dalton (Greenaway 1966), and the structural formulae of organic chemistry represented its major success. However, the Zeitgeist of the late nineteenth century was anti-atomistic. The growth of thermodynamic theories suggested that the whole of natural phenomena could be explained in terms of simple observations of energy and heat (a phenomenological excess comparable to that frequently found in the psychology in this century); this, as Bernal (1969) points out, 'in the hands of philosophers like Mach and chemists like Ostwald, seemed to promise an escape from the awkward materialism and radicalism of the atomic theory' (p. 590). (See Blackmore 1972.) This new positivism declared that matter and physical hypotheses such as atoms were no longer necessary, and that the whole of science could be deduced directly from elementary observations. Maxwell's kinetic theory of heat, indeed, implied the existence of atoms, but these were entirely hypothetical, and there was no direct evidence for the existence of atoms as measurable and countable material objects. Newton, of course, had been an atomist, but his mechanics, as generalized by Lagrange and Hamilton, lent itself to a picture of space in which properties varied only slightly from place to place, and this *field theory* type of looking at nature acquired great prestige from the Faraday-Maxwell development of the electromagnetic theory of light, leading later on to Einstein and the theory of relativity.

By the turn of the century, however, irrefutable evidence of the physical existence of atoms arrived, although even that was still debated by eminent physicists like Dumas;

the work of Perrin, Einstein (in his paper on Brownian movement) and others established the existence of atoms beyond any doubt – oddly enough at the same time as evidence was accruing, from the work of J.J. Thonsom and others, that atoms were not really 'uncuttable', but were made up of smaller particles, orbiting a nucleus in largely empty space. Later still came the agonizing questions about the wave or particle nature of these entities themselves, Heisenberg's principle of uncertainty, and the relation of quantum theory to relativity, but these are not really relevant to the point here made, namely that the 'existence' of a scientific concept, such as atom, remains debatable, in spite of its great success in mediating the explanation of old, and the acquisition of new knowledge; what is needed is the direct physical measurement of the concept in question. We may also note that in achieving such measurement, the concept itself may be changed in very fundamental ways, and that this transformation may give rise to many new and intractable problems.[1]

It always seemed likely that agreement on the 'existence' of intelligence would not be reached as long as the concept was based on essentially phenomenological evidence, however elaborate the statistical treatment; what was clearly needed was the demonstration of a physical basis for what before had been treated as a mentalistic phenomenon. The existence of such a physical basis was already implicit in the strong genetic determination of IQ measures (Eysenck 1973, 1979); as T.H. Huxley pointed out over a century ago: 'No psychosis without a neurosis', i.e. no mental event without a corresponding physical event underlying it. As will be shown later in this book, we now have a theory linking physiological mechanisms with IQ performance, and evidence that such a link is very close indeed; correlations in excess of .8 have been obtained on large random samples of the population between such tests as the WISC and our special theory-derived evoked potential EEG measures. Such results suggest that we have come quite close to the physiological measurement of the genotype underlying the phenotypic IQ test results on which we have had to rely so far. Furthermore, this new physiological index correlates almost equally highly with verbal and non-verbal tests (when corrected for attenuation), and its correlations with different subtests of the WISC are closely proportional to the loadings of these test on a general factor common to them all (Eysenck 1981b). Obviously these results are subject to replication, different interpretation, criticism, and general discussion; furthermore, many important types of experiment are suggested by this finding, both to establish it more firmly, and extend its coverage. But in principle it may be suggested that as the 'existence' of the atom was finally agreed only after its physical determination, so here also we may hope that the physical demarcation of intelligence will lead to a greater degree of agreement on its 'existence' than has been evident in the past.

In presenting in this book both the theory on which the new measure of intelligence is based, and the results of the application of this measure, we do not wish to imply that the success of the application necessarily proves the correctness of the theory. The theory is far-reaching and will require considerable support from direct experimentation, much of it in the realm of biochemistry and physiology. Even should it be found wanting, in part or in whole, this would in no way alter the empirical findings; from the psychological point of view the only direct link between theory and experiment is the hypothesis that there exists a probabili-

1 At present, for instance, it is believed by physicists that a proton is composed of three quarks, held together as entities called gluons pass between them. Evidence for such nuclear 'glue' came first in 1979 from collisions between high-energy electrons and positrons. Thus first the atom turned out not to be 'uncuttable', and to be composed of electrons, protons, positrons, etc., and now these in turn have proved not to be 'uncuttable' and to be composed of quarks, gluons, etc. Whether these entities in turn will prove 'uncuttable' remains to be seen.

ty, R, that a given message encoded in a series of pulse trains will arrive at its destination in the identical form in which it was encoded, while $1-R$ presents the probability of an error occurring during transmission. For longer sequences of recognitions, and assuming independence of probabilities, the probability of a longer chain of N events succeeding is simply: R^N. The theory then states that R is the basis of what we call intelligence, and goes on to make deductions about the direct measurement of R in terms of evoked potentials. Hendrickson's (1972) theory about the actual biochemical and physiological events concerning the transmission of neural impulses, and the occurrence of errors, is of course of major importance, but even should it turn out to be erroneous this would in no way lessen the interest and importance of the error-in-transmission hypothesis, and its link with the experimental determination of a physiological correlate of IQ (Hendrickson and Hendrickson 1980).

This formulation has many advantages. In the first place, it is truly causal; it tells us that higher mental functioning is dependent on the correct transmission of information, with the implication that correct transmission of information is a necessary and sufficient condition of successful cognition. This is the strong form of the hypothesis; a weak form (more cautious but less exciting) would simply state that accurate transmission of information is a necessary but not a sufficient cause of successful cognition. The data seem to favour the strong form of the hypothesis, as will be clear from a consideration of some figures. As pointed out before, the correlation between the Hendricksons' evoked potential measure (let us here denote it simply R) and Wechsler IQ is +.83. However, the variance of the Wechsler IQ is determined 80% by genetic, 20% by environmental causes; the latter are not, according to the theory, measured by R (Eysenck 1979, 1981b). When we take this into account, as well as attenuation due to unreliability (not large, but obviously not non-existent) the correlation suitably corrected commes fairly close to unity. Thus R and (purified) IQ seem to be almost identical, leaving very little room for non-R components of intelligence. The same conclusion is reached if we start with the general factor loading of R on the factor analysis of the matrix of intercorrelations between R and the WISC subtests; this amounts to .91, which, when suitably corrected, again leaves little variance over for non-R components of intelligence. (R does not, of course, cover what Burt called 'group factors', such as verbal, numerical, or perceptual ability as determined after extraction of a general factor; these presumably have different physiological indices. Possibly verbal and non-verbal abilities are characterized by different hemispherical functioning, a possibility which again could be investigated by means of the evoked potential.)

However clearly the indications may be pointing in that direction, the conclusion seems counter-intuitive. We tend to think of cognition in terms of problem solving which involves correct transmission of the elements constituting the problem, and perhaps the memory traces needed to create what Spearman called 'noegenetic' material, but firmly believe that the actual solution of the problem is different from, and involves separate mechanisms to those involved in the simple transmission of information. Sternberg, in his contribution to this book, has outlined his own development of Spearman's three laws of noegenesis, and has adduced much valuable experimental material to support these developments; it must be left to others to try and reconcile the experimental findings of Sternberg and the empirical results of the Hendricksons. At the moment there appears here to be a contradiction which requires solution, a contradiction perhaps as difficult to resolve as that between the wave and corpuscular theories of light. The nature of light is still an unresolved problem, with physicists thinking in wave terms on Mondays, Wednesdays, and Fridays, and in corpuscular terms on Tuesdays, Thursdays, and Saturdays – leaving Sundays to think about it in terms of both! Perhaps we shall have to follow their example.

The comparison of intelligence and atomic theory presents another interesting similarity. As was noted earlier on, the moment when atoms were universally recognized as fundamental building stones of the universe was also the moment when the conception of their nature was entirely transformed by the discovery that atoms were not uncuttable, but were in turn made up of a whole host of smaller entities – first electrons, protons, and neutrons, then neutrinos, and positrons and finally hundreds of wraith-like structures themselves (possibly) made up of quarks and other extremely odd and highly speculative elements. Much the same may be happening to the IQ. Usually regarded as a kind of entity, it clearly is no such thing; the apparent unitary nature of the concept is based largely on a rather elementary error to which most psychometrists will have to plead guilty. Practically all the correlational and factor analytic work on intelligence tests has used as its basic constituent the *test score*; correlations are nearly always between test scores, and factors report loadings of tests. Yet, as Eysenck pointed out long ago (Eysenck 1953, 1967b), identical scores can be achieved with entirely different sets of items solved correctly, incorrectly, or abandoned. In other words, identical test scores are often based on very inhomogeneous combinations of test items, and the very real possibility exists that by combining inhomogeneous items into total scores, much information is lost, and a completely artificial impression of homogeneity given. This line of argument suggests, first, that individual items should constitute the raw material from which correlations and factors are extracted, and second, that by failing to time for each item the duration of solution (whether right or wrong), or abandonment, we lose much important information. The writer suggested that such an analysis would result in breaking up the IQ into three major and probably largely independent parts, namely mental speed, persistence, and error.

This idea was followed up, first by Furneaux (1960), and later on by White (1973), whose contribution to this book constitutes the latest version of the latent trait analysis made along these lines. M. Berger has contributed a historical review of the development of these ideas, and there will be no attempt here to replicate this task. Let us merely note that the general outcome of the many tests of the general hypothesis of IQ inhomogeneity has been positive in almost every instance; there seems little doubt that like the atom, so the IQ also can be broken up into separate constitutents which require separate measurement. It is interesting to note that when we correlate these independent constituents with total IQ scores, it is the error component which is most prominent, as one might have anticipated from the elaboration of R as the major (possibly the only) cause of IQ differences. Much, of course, remains to be done; it will be particularly interesting to note the degree to which R, as measured psychophysiologically, correlates with mental speed, persistence, and error measured separately. Such an experiment has not yet been done, but it should throw much-needed light on the interrelations between these two hitherto rather isolated modes of thinking about problem solving and intelligence.

The theory of R as basic to intelligence also throws much light on the close relation observed between reaction time measurement (RT) and intelligence, and between inspection time (IT) measurement and intelligence. The two chapters by Jensen and Brand recount the facts; here we may be permitted to add a few words concerning the integration of the findings with the theory underlying R. Jensen notes that the closest correlation between IQ and RT is with RT variability; the brighter a given subject on IQ tests, the less is the variance of his RT scores. This is of course precisely what would be predicted on the basic of the theory developed by the Hendricksons. If the best RTs of bright and dull are roughly equal (as seems to be the case when the quickest reactions are being compared), then those with low R scores would have many long or very long RTs because of errors in the transmission of information through the cortex; it is these which lengthen mean RT, and produce an increase in variance over trials. The same theory can easily explain differences in mean RT between high and low IQ subjects, and the increase in correlation between IQ and RT with increase in the number of alternative stimuli.

Much the same may be said about the correlation between inspection time and IQ discussed by Brand. This type of measure, first used by Burt (1909) with very good results (he reports correlations in excess of .8) and lately extensively employed by Lehrl (1980) and Lehrl et al. (1975, 1980) with similar success, is of course a measure of speed of reaction, but may be reinterpreted in terms of R. Clearly, correct estimates of briefly presented figures are dependent crucially on correct processing of information through the sensory channels; errors require repetition and this redundancy slows up final decisions. Thus work on inspection time can be brought under the same umbrella as RT work in general, and variance on RT tasks in particular. This interconnection between the psychophysiological work of Hendrickson, the RT work of Jensen, and the inspection time work of Brand strongly supports the acceptance of R as a fundamental variable in psychology, and the major (perhaps the only non-cultural) constituent of IQ.

Let us now consider what is the major import of the empirical work here discussed. The major finding is that, along several independent lines (psychophysiological recording of AEPs, RT measurement, inspection time determination), IQ correlates very highly (.8 and above, without correction for attenuation) with tests which are essentially so simple, or even directly physiological, that they can hardly be considered *cognitive* in the accepted sense. There is certainly no question of problem solving; only simple sense impressions are involved, and simple motor movements in response. There is little or no room either for educational, cultural, or social status effects to arise; no subject taking part in these investigations would have the slightest difficulty in making the appropriate response given sufficient time. Thus we arrive at the astonishing conclusion that the best tests of individual differences in cognitive ability are non-cognitive in nature! This conclusion is certainly counter-intuitive, but it is difficult to see how it can be avoided, in the light of the evidence. (See also Table 1 in Chap. 9.)

We may consider the usual definitions of intelligence in relation to this finding. Intelligence is often described or defined in terms of success in problem solving, ability to learn, capacity for producing noegenetic solutions, understanding of complex instructions, or simply all-round cognitive ability. All these definitions or descriptions imply a properly functioning CNS in which errors in transmission are infrequent. This is a basic ability without which it is doubtful if any of these functions could be said to be capable of operating normally. It is astonishing, as already pointed out, that there seems to be little or nothing in IQ beyond simple information processing without error, but perhaps we should not be too surprised (Spiegel and Bryant 1978, Hunt 1980). There is the ever-present danger of homunculus thinking, i.e. the notion that somewhere in the CNS there sits a tiny man-like creature which absorbs the input and produces the output – whether problem solutions, IQ test answers, appropriate correlates, or whatnot. The actual solution process is certainly well outside our conscious experience; solutions seem to come to us from the depths of our brain, out of what, were it not for Freud, we might call the unconscious; psychology has certainly done spectacularly little to explicate the events involved. If our results are replicable, they would seem to suggest that there may be nothing more than the bringing together of sensory impressions with relevant memories, with the solution emerging with a facility depending on R!

Such a theory would certainly be regarded as reductionist with a vengeance, and of course materialistic and reductionist theories are very much out of favour in psychology at the moment. (This reminds one again of the degrees to which atomistic theories were out of favour in the latter half of the nineteenth century, with idealism setting the scene, and inspiring the Zeitgeist). With 'cognitive psychology' in vogue again, biologically and genetically oriented theories, harking back to the evolutionary development of mankind, find it hard to get a hearing, but the facts here published are too clear cut to be neglected, even by those whose Weltanschauung has little room for reductionism. Much, of course, remains to be done; what is reported here is only the beginning of a long climb out of the morass of mentalism to the safe refuge of a proper biological theory. But the signs are not unfavourable; the surprisingly high relations between IQ and R, RT, and IT require an explanation, even if that tentatively offered here be rejected. Clearly 'cognitive' theories of the usual kind do not furnish any sort of explanation; something new is required, and it will be interesting to see in what way traditional psychology will attempt to incorporate this new material. Whatever the outcome, it seems certain that a Kuhnian revolution has started in this field, and that entirely new concepts are required to accommodate the new evidence.

References

Bernal JD (1969) Science in history, vols 1–4. Watts, London

Blackmore JT (1972) Ernst Mach: his life, work and influence. University of California Press, Berkeley London

Burt C (1909) Experimental tests of general intelligence. Br J Psychol 3:94–177

Cronbach LJ (1957) The two disciplines of scientific psychology. Am Psychol 12:671–684

Eysenck HJ (1953) Uses and abuses of psychology. Penguin, Harmondsworth, Middlesex

Eysenck HJ (1967a) The biological basis of personality. CC Thomas, Springfield

Eysenck HJ (1967b) Intelligence assessment: a theoretical and experimental approach. Br J Educ Psychol 37:81–98

Eysenck HJ (ed) (1973) The measurement of intelligence. MTP Press, Lancaster

Eysenck HJ (1979) The structure and measurement of intelligence. Springer, Berlin Heidelberg New York

Eysenck HJ (1981a) A model for personality. Springer, Berlin Heidelberg New York

Eysenck HJ (1981b) Recent advances in the cross-cultural study of personality. In: Spielberger C, Butcher J (eds) Advances in personality assessment. Erlbaum, London

Furneaux D (1960) Intellectual abilities and problem-solving behaviour. In: Eysenck HJ (ed) Handbook of abnormal psychology. Pitman, London (reprinted in Eysenck 1973)

Greenaway F (1966) John Dalton and the atom. Heinemann Educational Books, London

Guilford JP (1967) The nature of human intelligence. McGraw-Hill, New York

Hendrickson AE (1972) An integrated molar/molecular model of the brain. Psychol Rep 30:343–368

Hendrickson DE, Hendrickson AE (1980) The biological basis of individual differences in intelligence. Pers Individ Diff 1:3–34

Hunt E (1980) Intelligence as an information-processing concept. Br J Psychol 71:449–474

Kamin L (1981) In: Eysenck HJ, Kamin L (eds) The intelligence controversy. Wiley, New York, pp

Lehrl S (1980) Subjectives Zeitquant als missing link zwischen Intelligenzpsychologie und Neuropsychologie? Grundlegenstud Kyberketik Geisteswiss 21:107–116

Lehrl S, Straub B, Straub R (1975) Informationspsychologische Elementarbausteine der Intelligenz. Grundlagenstud. Kybernetik Geisteswiss 16:41–50

Lehrl S, Gallwitz A, Blaham L (1980) Kurztest für allgemeine Intelligenz KAI: Manual Vless, Munich

Revenstorff D (1978) Vom unsinnigen Aufwand. Arch Psychol (Frankf) 130:1–36

Spearman C (1927) The abilities of man. Macmillan, London

Spiegel MR, Bryant, ND (1978) Is speed of processing information related to intelligence and achievement? J Educ Psychol 70:904–910

Thorndike EL (1927) The measurement of intelligence. Columbia University Press, New York

White PO (1973) Individual differences in speed, accuracy and persistence: a mathematical model for problem-solving. In: Eysenck HJ (ed) The measurement of intelligence. MTP Press, Lancaster, pp 246–260

A Splitting the Atom
The Break-up of the IQ

2 The 'Scientific Approach' to Intelligence: An Overview of Its History with Special Reference to Mental Speed

M. Berger

Introduction

> 'And does the history of mental testing belong in the history of experimental psychology? Not really......'
>
> Boring 1957

At the turn of the century mental testing was in a state of crisis, at least in so far as the psychology of intelligence was concerned. This aura of crisis is well illustrated by Spearman's (1904b) review of the early attempts to relate mental test scores to various criteria of academic competence.

> Thus far, it must be confessed, the outlook is anything but cheerful for experimental psychology in general. There is scarcely one positive conclusion concerning the correlation between mental tests and independent practical estimates that has not been with equal force flatly contradicted; and amid this discordance, there is a continually waxing inclination – especially among the most capable workers and exact results – absolutely to deny any such correlation at all. (Spearman 1904b)

The early crisis was helped towards resolution by a number of important papers. Two were published by Spearman (1904a, b) and one by Binet and Simon in 1905. Together, these papers had a number of ramifications, perhaps the most important of which were the foundations that were laid for applied intelligence testing, for psychometrics, and for the various theories of intelligence asociated with correlational psychology.

With certain exceptions, notably Binet and Ebbinghaus, most of the tests used in the early period of mental testing were single, homogeneous tests directed mainly at measuring sensory and motor processes and memory. These tests were derived from experimental psychology laboratories and were used in the study of inheritance (exemplified in the work of Galton), in the prediction of academic success, or in other aspects of research on individual differences (Boring 1957, Freeman 1939).

The tests used by Ebbinghaus and those being developed by Binet were aimed at the 'higher mental processes'. By 1898, according to Boring (1957), it was possible to show that in so far as item content was concerned, the Binet-type item had 'won out'.

In this early period, the tests were not organized into scales, and if a number of them were administered concurrently, the scores were not combined (Freeman 1939). The development of a scale which produced a total score is generally attributed to Binet and Simon. However, as Wolf (1973) indicates in her biography of Alfred Binet, the idea probably originated in the set of tests used by Blin and Damaye for the diagnosis of subnormality. Their tests, produced before the 1905 scale, yielded a total score and were sufficiently developed to provide a crude range indicating 'normal' performance.

The development of the first scale of intelligence was not a 'fortuitous event' (Wolf 1973). For over 2 decades, Binet had been searching for a procedure which would enable a discrimination to be made between the then accepted medical classification of three grades of mental deficiency. The central idea of the 1905 scale published by Binet and Simon was that individuals of different degrees of intelligence at the same level of maturity could be distinguished by the number of tests passed. The component tests of this scale were graded according to difficulty level and the notion of an age scale was already present (Freeman 1939).

Taken separately, neither the content of the items nor the idea of a scale was novel. The novelty was to be found in their combination and in the provision of explicit instructions for administration and scoring.

On the 1908 scale and later editions of the Binet, mental age was computed by first finding the basal age and then adding 1 year for each five items passed above that level. This procedure of adding together disparate items to produce an agglomerate score is of paramount importance because it represents an approach to quantification which has affected the scoring of psychological tests ever since. As Du Bois (1970) points out

> What had been a 'test' now became a sub-test or item of a scale which as a totality yielded a composite measurement of a complex function. (p. 36)

Further early developments served to consolidate some of the characteristics of early tests. The first point scale was developed by Yerkes, Bridges, and Hardwick. Instead of mixing items at one age level, it grouped together items of homogeneous content graded in difficulty within subtests. The total score was then related to a table of norms from which a mental age was obtained. In principle, it did not differ from the Binet tests.

A second important development, group tests, also occurred in this early period. Although Otis is credited with this 'invention', there is some controversy as to who was responsible (Linden and Linden 1968). As an innovation, group testing facilitated data collection, an advantage that was fully exploited during the First World War (Boring 1957, Freeman 1939, Du Bois 1970).

There is little doubt that by the end of the First World War intelligence testing as we know it today had become established as the dominant technique in the psychology of intelligence. Spearman (1904b), while contributing to this trend through his advocacy of the technique of correlation, refused to associate himself with this approach to intelligence. Thus, after reviewing the work of Binet and Henri which had been published in 1895, he commented on an unacceptable 'new feature' in their tests:

> Hitherto, these had been of the most elementary and unequivocal nature possible, as befits the rigour of scientific work.... Binet and Henri appear now to seek tests of a more intermediate character, sacrificing much of the elementariness, but gaining greatly in approximation to the events of ordinary life. The result would seem likely to have more practical than theoretical value.

In the 1890s mental testing was an integral part of experimental psychology. Twenty years later it had established a separate identity, which it still retains. Yet even as the mental test tradition was emerging, there were psychologists (Spearman, Thurstone, McFarland, and Peake and Boring among others – see below) who recognized the dangers of such separation, who refused to identify themselves with the mental test movement, and who actively maintained an allegiance to experimental or scientific psychology as they understood it. The prime concern of these proponents of a 'scientific approach', was to elucidate the nature of intelligence within a scientific framework, and the new tests diverted psychology from this task. The new tests provided a particular style of item and structure for tests as well as procedures for constructing and evaluating tests and a stimulus to use the total score as the main unit. From the point of view of the experimentalist, these features served to confuse rather than clarify scientific analysis. Some of the objections are considered in the following paragraphs.

Some Inadequacies of Tests of Intelligence

The inadequacies of intelligence tests have been commented on periodically. Thorndike et al. (1927), while recognizing that the then current tests were an improvement on their predecessors, nevertheless recognized three sources of deficiency. According to their view, the tests were ambiguous in content, their units arbitrary, and the results of uncertain significance. In 1925 Thorndike suggested that ability should be analysed into level, range, and speed. It was acknowledged that for practical purposes, a test battery which combined speed, level, and range (or extent) in unknown amounts might well be useful. However,

> For rigorous measurements... it seems desirable to treat these three factors separately, and to know the exact amount of weight given to each when we combine them. (Thorndike et al. 1927, p. 25)

For Thorndike the measurement of extent and speed posed no problem. The former could be accomplished by simply counting the number of tasks correct in a sample of problems from all areas of intellectual functioning. Speed too could be readily measured. The measurement of difficulty, however, posed the greatest problem. If difficulty could be properly measured, 'level' would readily succumb to measurement, and it would then also be possible to quantify 'extent' and 'speed' at any level. Hence, for Thorndike, the quantification of difficulty represented the central task in the measurement of intelligence. This problem will be considered later.

In 1927 Spearman criticized the 'hotchpotch' procedure employed in scoring intelligence tests. What he failed to see was the relevance of his comments for the tests of homogeneous content which he used. Spearman's concern was the crudity of the procedure inherited from Binet, 'the prevalent procedure of throwing a miscellaneous collection of tests indiscriminately into a single pool....' (p. 71). Spearman, like Thorndike, also recognized the need, as part of a scientific approach, to 'dissect' the subject matter under investigation. The problems entailed in the use of an agglomerate score, and the need for 'dissection' have been reemphasized in the more recent critical analyses of intelligence tests put forward by Furneaux (1961) and Eysenck (1967, 1973).

In a paper published in 1967 Eysenck presented a number of major criticisms of contemporary approaches to the study of intelligence. This paper brought together the theoretical arguments and empirical evidence for his view that psychometrics had 'become almost completely divorced' from the mainstream of experimental psychology, that research was overly dependent on factor analysis, which could not on its own provide the answers that psychologists sought in their study of intelligence: further that the basis of intelligence measurement, the total test score, confounded a number of components and was thus unsuitable as the basic unit for the analysis of test performance. Finally, he noted the lack of concern with 'mental speed' and suggested that it should be 'restored to its theoretical preeminence as the main cognitive determinant of mental test scoring ability'.

The fundamental criticism advanced by Eysenck concerns the assumption that equal scores on a test are intellectually equivalent. According to common usage, it is assumed that if individuals obtain the same total score on a test, then such scores have equivalent psychological significance.

Eysenck (1967) illustrates the problems in the following way. For any given test item, the possible outcomes are 'correct', 'incorrect', 'abandoned', and 'not attempted'. If each individual obtains the same total score, and provided this score is below the maximum, it is possible that they have done so using quite different routes: person A gets two correct and fails the remaining three items; Person B also gets two correct (different to those of A), gets one wrong, decides not to attempt one, and abandons the remaining item. Even this diversity of pattern is oversimplified. Any correct solution could be a 'lucky guess' and one that is correct a 'mistake', which might be uncovered by close questioning of the subject. Eysenck (1967) asks 'Can it really be maintained that the mental processes and abilities... are identical merely because they obtained the same final mark?'

Thorndike and his co-workers were also aware of this problem, although they discerned it in a somewhat different context, that of the measurement of speed.

> In the measurements that are actually used it is customary to have the time a mixture of (1) the time spent in doing some tasks correctly, (2) the time spent in doing other tasks incorrectly and (3) the time spent in inspecting other tasks and deciding not to attempt them. This confusion may be permissable, or even advantageous, in the practical work of obtaining a rough measure of intellect at small expense of labour and skill, but for theory at present and for possible improvement of practice in the future we need to separate the speed of successes from the speed of failures. (p. 33)

As Eysenck (1973) notes, Thorndike never followed through the implications of his critique. Instead, it was Furneaux (1961) who attempted to incorporate these as part of his procedure for the analysis of item solutions. This involved replacing 'inspection time' by 'the time spent in attempting a task and deciding to give up attempts at solution' (Eysenck 1973, p. 191).

The need for a detailed conceptual analysis of item solution as a prerequisite of a scientific approach was a theme in the writings of both Spearman and Thorndike. This theme was taken further by Furneaux (1961). Like his predecessors, Furneaux accepted that the empirically established relationships between IQ and other criteria justified the practice of crude testing. What remained as unsatisfactory was the incompleteness of the ensuing description of test-taking behaviour. Furneaux argued that the success of applied testing may simply be a consequence of the tests and the real-life situation reflecting an even closer relationship between only certain aspects of test-taking and real-life performance. The remaining components of the test score then simply act as a source of error. For Furneaux, the appropriate scientific goal is to minimize the error, by a process of maximizing the number of categories into which test-taking behaviour can be classified and ultimately scored. These categories should be refined to the point that further subdivision is no longer possible. Hence, for Furneaux:

> ... the only really satisfactory approach to the study of test-taking behaviour is that of the thorough-going logical atomist.

While he never formulated it as such, what Furneaux appeared to be aiming at was a fractionation of the observed score variance into its component sources of variance, a model compatible with 'the analysis of variation'. The major constituents to emerge in Furneaux's model were speed, accuracy, and continuance.

Spearman's (1904b) prognostication – that the Binet-Simon style tests were 'likely to have more practical than theoretical value' – appears not to have been falsified. The Binet test has undergone several revisions (Terman and Merrill 1960) and together with a number of tests modelled on the same pattern, continues to be used for practical but increasingly questionable purposes in the various fields of applied psychology. Their theoretical significance is questionable even though a number of test authors have made excursions into debates on the nature of intelligence (e.g. Wechsler 1958). The mental test tradition, as this stream may be called, also played an important part in the development of test theory (Gulliksen, 1950, Lord and Novick 1968) and for this reason as well is of significance.

The proponents of a scientific approach have also had to contend with the second major stream in the psychology of intelligence, that encompassed by factorial analysis of intelligence tests and factorial conceptions of ability.

If we follow Boring's (1957) historical analyses, this second main stream of development must find its source in Darwin's *Origin of the Species*, published in 1859. Ten years later, Galton's *Hereditary Genius* appeared, followed in 1883 by *Inquiries Into Human Faculty and its Development*. This last-mentioned book by Galton 'has sometimes been regarded as the beginning of scientific individual psychology and of mental tests' (Boring 1957, p. 483).

The major technical innovation was the invention of the technique of correlation by Galton and its refinement by Pearson. It was Pearson who also invented a technique of factor analysis, later rediscovered by Thurstone (Burt 1949). Factor analysis, according to Burt, was used like other mathematical calculations 'merely as aids to verifying... hypotheses which they had already reached on broader grounds'.

The philosophical conceptions of mind and intelligence, as they had evolved by the end of the nineteenth century can be broadly divided into those which regarded mental processes in unitary terms, those which regarded mind as a set of faculties, and those which conceived of the intellect as being hierarchically ordered, the latter an eclectic view advocated by McDougall. Historically, and to some extent today, the faculty view finds its major proponents in North America whereas the hierarchial conception, with its lineage traceable through McDougall, Burt, and Vernon is more common in the U.K. (Vernon 1961).

Spearman's position, particularly in its early expression, was somewhat anomalous. In his early writings he was an advocate of what he called the 'theorem of intellective unity'. He adopted the terms 'general ability' and 'general intellectual power' directly from Galton's wirting on inheritance, as he did the terms 'special aptitudes' or 'special powers'. In his later years, as is generally known, he accepted more explicitly the hierarchical structure (Spearman and Jones 1950). In this sense, his views can be aligned with those who adopt a factorial conception of intelligence.

Spearman, as already noted, did not regard the Binet tests as scientifically respectable (Spearman 1904b), and continued to employ tests of homogeneous content. In a second paper (Spearman 1904a), he also advocated the use of correlational techniques, suggesting that these might have the important consequence of being able to reveal the '*hidden underlying cause of the variations*'. While both Spearman and factor analysis begin with correlations, technically they soon part company. As Burt (1949) has pointed out, Spearman only partially accepted the work of Galton and Pearson because he believed that there was a basic difference between the physical measurements used by Galton and Pearson and psychological measurements, the latter being significantly more error prone. It is possibly for this reason that he devised the rank-difference technique and the correction for attenuation. Further, as Burt states 'the novel feature in his procedure consisted not in 'factor analysis' as now understood, but rather in a method which has a close affinity with so-called 'canonical analysis'. Spearman himself made little use of the word 'factor': it is as Burt (1949) states 'scarcely mentioned'.

Factor analysis has had a significant influence on the study of intelligence. This influence persists (Cattell 1971, Guilford 1967, Guilford and Hoepfner 1971) and may well be extended even further with the availability of computer-based data analysis. Whether or not it can be regarded as part of experimental psychology is an open question. Boring (1957), while recognizing its technological importance, does not consider it appropriate to include factor analysis and the theories it generates as part of the history of experimental psychology (see p. 481).

The important limitations of factor analysis are now well known (Butcher 1968, Heim 1970) as are the limitations which emerge when attempts are made to interpret the results of factorial studies (Vernon 1961). While some factorial techniques are likely to remove some of the significant technical difficulties (Lawley and Maxwell 1963), a number of problems will remain, particularly with regard to already published studies.

Furneaux (1961) has pointed out some of the problems of interpretation that accompany factorial solutions. He suggests, for example, that apparently established group factors may well be artefacts. In examining data from Thurstone's Primary Mental Abilities, reanalysed by Eysenck (1939), Furneaux suggests that different interpretations of the data are possible. Whereas the original analysis suggested a differentiation between visuospatial and artithmetical tests, an alternative explanation could be that the differences arose because the tests defining these factors differed in the extent to which they measured speed and accuracy.

Criticisms of this type can of course be applied to all of the research on intelligence, and particularly to the multitude of studies which base themselves on factor analysis. The validity of such criticisms is, however, very much dependent on the empirical and theoretical status of the alternative framework. If 'speed' and 'accuracy' are psychologically, rather than only linguistically, meaningful concepts, then, as Furneaux suggests, most of the factorial work 'will eventually have to be repeated'.

Eysenck's critique of factorial approaches is directed partly at the use of crude scores – the 'hotchpotch'; at the way in which factorial approaches have become 'divorced from both psychological theory and experiment'; and at the failure to recognize that despite its usefulness as a tool, the technique cannot cope with the various fundamental demands placed on it.

Among the critics of conventional testing and factor analytic traditions, there is a lineage with several characteristics. Not only do they espouse a 'scientific' approach but also advocate the single test item as the basic unit for the analysis of test performance. An outstanding feature of this group is the emphasis which is given to the role of 'speed' in test performance. All three characteristics are illustrated by the following statements:

> These more recent researches have been conducted, for the most part, by the individual method of timing and have conformed to scientific testing procedure. (McFarland 1928)
>
> ... speed is so important in the intelligent act that it has seemed to us the first factor to be studied if we are ultimately to come at an understanding of the nature of intelligence. (Peak and Boring 1926)

Relatively few studies have been carried out within the framework of this approach, and there have been even fewer attempts to develop a substantial theoretical framework to encompass such research. Among the outstanding studies are those of Peak and Boring (1926), Sutherland (1934), Slater (1938), Tate (1948), Cane and Horn (1951), Furneaux (1961), Russell (1968), and Brierley (1969). Some attempts at theorizing are apparent in the work of Thorndike (1925), Thorndike et al. (1927), and Thurstone (1937), but the most substantial analysis is that developed by Furneaux (1961). His work has also influenced a number of subsequent approaches, notably those of White (1973a, b) and Van der Ven (1971, 1974) among others.

Mental Speed: An Introduction to Theory and Research

The investigation of individual differences in the timing of stellar transits – the personal equation – probably represents the earliest of the systematic attempts to examine speed in human abilities (Boring 1957). Galton's interest in individual differences encompassed speed of reaction and in one form or another, psychologists have retained an interest in the speed of mental functioning up until the present day. This is witnessed by the ongoing research on speed of reaction in the elderly (Botwinick 1973), on temperamental differences in speed (Eysenck 1967, Cattell 1971), and on reaction times (Laming 1968, Smith 1968).

Before 1900, much of the research on speed was focused on individual differences in reaction time and the factors which influenced reaction speed. After 1900, there emerged a newer trend that began to focus on speed in relation to mental ability. As McFarland (1928) points out in his review of research on the role of speed in mental ability, the emphasis shifted to investigations of the relationship between 'quickness' (as measured by reaction time) and 'brightness', indexed by school performance and teachers judgements. The study by Burt (1909) in which he correlated card sorting and alphabet sorting speeds with headmasters judgements of intelligence typifies the type of research carried out.

With the development of tests of intelligence, a new criterion of 'brightness' became available to researchers, and it was with such criteria that relationships with reaction time were sought. This research has persisted although its focus is changed. Much of the current research on reaction time and intelligence is concerned with aging, reaction time being regarded as the major index of speed decline in the elderly (Botwinick 1973). Earlier studies were concerned mainly with younger subjects, and the problems they investigated had a different theoretical orientation.

In 1916, an important conceptual complication was introduced by McCall (quoted in McFarland 1928) when he raised the question of the relationship between 'speed tests', 'power tests', and intelligence. As his criteria of intelligence, McCall employed teachers ratings, school marks, and composite test scores. The speed tests included cancellation and addition tasks, and vocabulary and sentence completion were used as tests of power. This study is historically important because it was probably the first to introduce the speed/power distinction.

With the advent of group tests of intelligence and particularly their widespread use during the First World War, psychologists and testees alike became concerned with the effects of time limits (May 1921; see McFarland 1928; Ruch and Koerth 1923). Although various studies found substantial correlations between limited and unlimited-time scores, they tended to be methodologically inadequate. Hunsicker (1925) for example pointed out that because of the wide range in abilities, a single test such as Army Alpha was a speed test for the more intelligent and a power test for the less able of the subjects.

By 1925, research had become more sophisticated and a more differentiated conceptual scheme had emerged. Hunsicker (1925) commented on the need to control for accuracy, and she introduced the use of individual item-times. Also in 1925 Thorndike proposed a scheme whereby intelligence was conceptualized in terms of level or power, range, and speed. By 1925 as well the role of persistence was implicated in test performance and Spearman's questioning the existence of a separate speed factor had entered the literature.

Thorndike and his colleagues (1927) were responsible for one of the major theoretical analyses, in which they posited various dimensions of the intellect, among which was 'speed'. Their analysis had an important impact on a number of researchers (e.g. Peak and Boring 1926) both at the time of it's exposition and subsequently (Furneaux 1961, Eysenck 1973). Several other important theoretical analyses and empirical studies appeared in the following decade (Spearman 1927, Thurstone 1928).

The last of the major foci for studies of speed probably had its origins at the beginning of this century when speed as a special ability was being investigated (Tate 1948).

The nature of the problem of speed is illustrated by Hunsicker's (1925) comments, viz:

> The history of mental measurement shows few if any questions that have given rise to more general and persistent enquiry than has this one, the relation between rate and ability ... there is full agreement that there are individual differences in rate of work and individual differences in ability ... the relationship between these two variables is the crux of the disagreement. Is there *any* relationship? If any, how much? Is the quality of one trait revealed in the quality of the other? Is rate of work any indication of mental ability? If it is, of what significance is the fact for mental measurements?

Although the approach to these questions has changed since they were summarized by

Hunsicker (e.g. the use of factor analysis) the questions themselves have not changed. In one form or another, they have persisted through the years (Spearman 1927, Spearman and Jones 1950, Tate 1948, Lord 1956, Jones 1959, Brierley 1960, 1969; Cattell 1971, among others), and despite the substantial body of research, psychologists appear still to be confronted by 'this vexed question of speed' (Cattell 1971, p. 64).

Mental Speed: A Conceptual Digression

Apart from Peak and Boring (1926) and, many years later, Furneaux (1961), 'mental speed' has not been the subject of a careful conceptual and technical analysis. Rather, it has been treated as a sort of 'mini black box' slotted into a broader framework. This is particularly so in the multitude of factor analytic studies which identified various 'speed factors'. Indeed, the psychological research on speed is a prime exemplar of what Harré and Secord (1972) identify as 'the conceptual leap from the theory to the operation ... over-emphasizing empiricism at the expense of conceptualization' (p. 36).

In 1927, Spearman asserted

> As regards the measuring of speed, there is no great difficulty; for (with suitable arrangements) not much risk is run inferring the duration of a person's mental processes from the time he takes to respond to the stimulus. (p. 245)

This seemingly simple prescription repeated by Furneaux (1961) conceals a number of major problems: it presupposes that all the mental processes between 'stimulus and response' are directed at 'problem solving' and that there is a suitable procedure for measuring these. Neither of these suppositions is acceptable, for reasons which will be presented in this and the next section of this chapter.

Peak and Boring (1926) have described some of the possibilities that might account for a difference in the amount of time required by two subjects to solve a test item. As they state, 'the loss of time may be intersitial or it may be inherent in the intelligent act'. In the interstitial case, the time difference arises because the slower subject, while performing the relevant operations at the same speed as the fast subject, lost time 'by irrelevant activities or by self distraction'. In the alternative case, according to Peak and Boring, the time loss may 'be inherent, it can be found that the constituents of the act occur more slowly in the poor subject than in the good subject .... Such a localisation we think of as a first step toward a solution of the problem of the nature of intelligence'. On the basis of their investigation, unfortunately confined to only five subjects, Peak and Boring concluded in favour of the latter view. It is also unfortunate that they did not attempt to develop a theory to account for their findings. Until the appearance of Furneaux's (1961) analysis of problem solving, there was no significant attempt to develop such a theory.

Another of the findings of Peak and Boring, that of a significant correlation between item-solution time on a test of intelligence and a measure of reaction time, inspired a number of attempts at replication (Goodenough 1935, Farnsworth et al. 1927, Lemmon 1928), all of which failed to confirm their results. However, none of these other studies recognized the most important feature of the original, namely, that the index correlated with reaction time was based on individually timed items. Instead, they correlated time to complete the test, and total scores, with their own measures of reaction time.

The Peak and Boring study was published at a time when most psychologists appeared to have a fairly clear conception, at least at the operational level, of what constituted a test of speed. The conventional speed tests required subjects to engage in repetitive activities, such as letter cancellation, detecting differences in simple shapes, adding three digits, and so on (Burt 1909, McCall 1916, as quoted in McFarland 1928; Highsmith 1924, Hunsicker 1925). In 1943, Cattell defined speed as the 'rate of repetitive performance, where all content material is perceptually given, through all cognitive levels'. Thurstone (1937) regarded speed in terms of 'the number of tasks that are completed in unit time, and these tasks are usually easy'. A similar definition, emphasizing the easiness of tasks has been given by Anastasi (1968), and most of the published studies on speed employ tests which conform to these prescriptions (e.g. Mangan 1959, Lord 1956, Lohnes 1966). In the majority of studies, no attempt is made to time individual items and it is implicitly assumed that inter-item time is a legitimate component of 'mental speed'. While technical difficulties in item-time measurement are no doubt important when items are answered rapidly, this is not the case when more difficult tests (such as 'power' or 'level' tests) are used. However, even when such tests are used, it is usually 'total time to complete the test' or 'number of items solved on a time-limit test' that provides the index of speed.

A somewhat different approach to the measurement of 'mental speed', following the influence of Peak and Borin, is also evident in the research literature. A number of investigators have used individual item times in their studies (Sutherland 1934, Slater 1938, Tate 1948, Cane and Horn 1951, Furneaux 1961, Russell 1968, Brierley 1969). These studies have also employed items of the non-repetitive type at different levels of 'difficulty', either on their own or alongside conventional speed measures. Such a divergence in trends was noted by McFarland in 1928. The characteristics of these two distinct trends in research can be summarized in terms of differences between the timing procedures used, and the content and difficulty levels of the items. What they share is the concept and the problem of 'speed' measurement.

Item solution time, test completion time, or rate are gross measures. They span a sequence of events which may be different for the individuals being measured and yet lead to a conclusion that they have produced equivalent performances. A crude example will illustrate this. Individual 'A' completes an item in 10 s, as does individual 'B', except that the latter happened to break the point of his pencil and had to get another before he could record his responses. Individual 'B' actually took only 3 s to get the solution but spent the rest of the time exchanging pencils. Are we to conclude that the speed of B is equivalent to that of A?

While the above example can be dismissed as an instance of 'random error' in the time measurement, it is readily replaced by a more relevant psychological analysis, derived from the study of reaction times and the fine-grain analysis of certain motor acts such as tapping (Frith 1973, Spielman 1963).

In the study of reaction times, a number of investigators have been concerned to divide up the total time (T) into at least two components, the time occupied in executing the motor act and the time occupied by the 'mental events'. Birren (1964) has reported that the movement time in simple reactions is not appreciably altered (i.e. the muscular reaction) as individuals get older. The age effects on reaction time appear to be more a consequence of the other aspects of reacting. For present purposes, it is sufficient to note the subdivision of T into Movement Time (MT) and the Reaction Time (TR).

i.e. $T = MT + TR$

In his study of problem solving, Furneaux (1961) attempted to remove the equivalent of MT from the item-solution times by special measures taken before the problems were submitted to his subjects so that his time measures were in effect those quivalent to TR in the above equation. While such a refinement may seem of minor significance when dealing with events of extended duration, for example 2 or 3 min, MT or its equivalent may occupy a substantial proportion of T when the full sequence is 60 s or less.

The components of TR have been the subject of a number of conceptual analyses in studies of choice reaction time. For present purposes, it will be sufficient to describe that presented by Smith (1968) in his review of research on choice reaction times, a paradigm which is more appropriate to test-item solution than is simple reaction time.

In addition to MT, Smith (1968) identifies four components:

1. The raw stimulus is 'preprocessed' until a representation of it is formed.
2. The preprocessed stimulus is then compared with some other model in memory. On the basis of these comparisons, the stimulus is categorized.
3. An appropriate response is selected.
4. The response to be executed is programmed.

A theoretical analysis such as this is useful in helping to conceptualize the possible sequence of events, although even at this simple level, there are a number of problems. For example, theoreticians are as yet uncertain of which, if any, of these stages occurs in parallel with one or more of the others or whether the events are serial. Whatever the case may be, there seems to be agreement that choice reactions can be conceptualized in terms of a number of components each spanning some period of time. This type of approach has of course been used by a number of writers. Welford (1969), for example, has outlined the difficulty in deciding which of a number of components is primarily responsible for the observed slowing in the sensory motor performance of older people. In discussing the possible sources, he suggests that similar factors may be responsible for the slowing also observed in 'mental tasks'. Welford (1969) has listed the component processes as including:

1. Recovery of material from memory
2. Short-term retention
3. Strategies of action

On the basis of the foregoing, it would not be inappropriate to conceptualize TR comprising a number of components (c) which take time to occur, viz:

$$TR = C_1 + C_2 + C_3 + \cdots + C_n$$

so that, T now becomes

$$T = MT + C_1 + C_2 + C_3 + \cdots + C_n$$

The above 'equation' should not be regarded as an algebraic statement but as a shorthand psychological formulation.

The analysis of T can be taken at least one stage further. In problem solving, not all of the 'C' components need necessarily be involved in the mental processes 'doing' the actual solving. Component C_1 might be 'trying to understand what the problem is', C_8 might be the 'checking mechanism' suggested by Furneaux (1961). For purposes of discussion, let C_4 be that component in which the 'brain is working on the problem'. Now, possibly in the same way that vigilance cannot be sustained indefinitely, or that even in such a simple task as tapping a stylus on a metal plate, there are gaps in performance (Frith 1973), it is probable that there will be 'gaps' or 'blocks' (Bills 1931) which arise during C_4. These may arise because of something analogous to work-produced fatigue, distractions, some microrhythm of the type described by Wolff (1967), or even the sort of mechanism which governs intermittent visual fixations (Shackel 1967). On the further assumption that we are able to discriminate and hence to time the 'work' and 'rest' 'interrupts' in C_4, we would be able to express C_4 as made up of a number of components. Thus

$$C_4 = W_1 + B_1 + W_2 + B_2 + \cdots + W_n + B_n$$

where W_1, W_2, etc = the times spent on actual work on problem solving and B_1, B_2, etc = the interrupt times.

Hence T can now be rewritten as

$$\begin{aligned} T = {} & MT + C_1 + C_2 + C_3 \\ & + (W_1 + B_1 + W_2 + B_2 + \cdots + W_n + B_n) \\ & + C_5 + \ldots + C_n \end{aligned}$$

The magnitude of T will also be a function of some effect of the 'perceptual complexity' of the problem. To express this, it may be necessary to introduce a constant multiplier for some of the C components, or to add a constant to others. Different constants may be required for yet other C components depending on the difficulty of the item, previous experience with tasks of this type, and so on.

The main point to arise from this excursion into psychological atomism is 'Which of these component times (one, some, or all) is to provide the index of speed?'

In his 1967 paper, Eysenck presented a table illustrating this question for total score. A similar table has been drawn up to illustrate the same point with reference to 'speed'.

Table 1. An illustration of how hypothetical time components could summate to produce identical solution times, despite the components themselves being different

Subject	Component				Solution time
	C_1	C_2	W	B	
1	2	2	5	1	10
2	1	1	6	2	10
3	1	5	3	1	10

In the above table, only some of the C component times have been utilized and the W and B component times have been summed.

The atomization of T has been confined to a general statement about an individual item time. No attempt has been made to include the additional components that might arise when the total time (TT) to complete a test is being considered. In this case TT would comprise T's for each item as well as the times taken between items. Also, it is possible that certain of the components of T might not be called into operation once the subject has had some practice at solving problems of a certain type.

The analysis of problem solving processes is not yet sufficiently far advanced for any pronouncement to be made on the validity of the foregoing suggestions. There does, however, appear to be some agreement that a multicomponent model is relevant to complex motor performances and to choice reaction times (Welford 1969, Smith 1968). There is further a suggestion that such a model is appropriate for mental tasks (Welford 1969). Furneaux's (1961) speculations about the nature of the problem solving mechanism would also support this contention. Similarly, some of the models described by Newell and Simon (1972) are consistent with such a formulation (e.g. information processing systems (IPS) which involve multicomponent processes).

Thus, while a multicomponent model seems to be appropriate, it is difficult to specify which of the many possible components is to be regarded as providing a basis for speed measurement. It should be emphasized that the preceding discussion has been mainly concerned with the time to solve the single item, an that even at this level, there are a number of complications in conceptualizing 'speed'. These complications are extended when the unit becomes 'the number of items completed or solved' in a given period or some transformation of this index. There is as well a major constraint which resides in the technology and data gathering procedures used. These are considered shortly. No attempt has, or will be made, to speculate on the mental structures or processes, electrochemical or otherwise, which 'take time' and which occupy the 'W_1', W_2', etc. segments of problem solving or information processing, or on any of the other components. Further, one may readily question the basic approach adopted, based as it is on an underlying computer model and its mechanisitic implications. Nevertheless, it should at least be apparent from the foregoing that mental speed is itself a complex action which deserves much more attention than it has received in the past.

Given the undoubted lack of clarity in conceptualizing the central, cognitive, or mental notion of speed, it is not surprising that further confusion abounds at the empirical level, that of speed measurement.

Mental Speed: A Digression into Measurement

The inconsistencies and contradictions in the voluminous literature suggest that either speed is an unstable dimension of intellect or that inappropriate units and methods of measuring it have been frequently employed. (Tate 1950)

The procedures for measuring 'speed' can be classified into those which time solutions to individual items and those which record the time to complete a block of items. In the latter case, researchers then proceed to derive some index of rate, exemplified by average item solution time, total time to complete '*n*' items, number of items completed in '*n*' minutes, etc. Either type of index can be obtained from group or individual testing situations, and in some studies (e.g. Peak and Boring 1926) mixed procedures were used.

In this section, it will be argued that none of these procedures provides an adequate basis for the measurement of speed. This will be followed by an attempt to link measurement to theory.

A fundamental problem is of course the assumption, most often implied rather than supported, that there is an isomorphism, or at least a very close approximation, between test performance and the underlying 'mental speed'. Apart from the technical difficulties which will be considered in this section, the measurement process is inevitably linked to the way in which speed is conceptualized. Here again, the relationship between measurement and theory is at best trivial in the majority of studies. It is argued that until such time as a full theoretical and conceptual analysis is forthcoming, much of the research will be little more than empty empiricism. In this section, some attempt will be made to relate the problems of measurement to the earlier speculative model of mental speed.

Group-administered tests obviously facilitate an efficient collection of data, but group testing has a number of inherent problems which adversely influence the quality of data collected. Some of these problems are especially important in speed measurement irrespective of the type of index ultimately used in the data analysis. Other problems differentially affect individual item and 'block of items' times. Supporting evidence for these statements is somewhat limited, mainly because authors generally fail to comment on limitations in published studies. However, the nature of the difficulties can be assessed from a number of reports (Hunsicker 1925, Tate 1950, Lord 1956, Lohnes 1966, Russell 1968, among others), as well as from more general analyses of group testing (Heim 1970, Vernon 1960, Cronbach 1970, Anastasi 1968).

Hunsicker (1925) employed both group and individual data collection procedures. In describing her study, she noted that even though the groups were in the process of being tested, they were 'in the main, although not entirely free from interruption'. She also expressed her concern with the 'dishonesty' which arises in a group test setting, citing as evidence one (unreferenced) study which revealed that 'fifty per cent of the class had cheated'. The Hartshorne and May (1928) studies clearly indicated the severity of this problem on even simple 'speed' tests. As they noted:

> ... even such slight changes in the situation as between crossing out A's and putting dots in squares are sufficient to alter the amount of deception both in individuals and in groups. (p. 382)

This problem is not confined to children or to era. In his 1956 study, Lord pointed out that for one of his measures of speed – viz. the number of the last item attempted – there is '... reason to believe that many or all of the examinees who answered the last item of the speeded tests skipped many items or responded at random', despite being instructed not to do so.

Hunsicker (1925) also cites evidence for the unreliability of data obtained from group testing. After assessing the various procedures for collecting group data, she states

> Not one has been found which gave evidence when in actual use, of any fair degree of control or elimination of irrelevant factors. In all likelihood, group testing by its very nature increases not only the number but the effect of disturbing elements in the situation.

After comparing her group and individual data, she states further '... the conclusion seems beyond cavil that the group method is not dependable for securing measures of rate'. As a consequence of the problems encountered, Hunsicker discarded her group data.

A more recent example of such difficulties is provided by data from Project Talent (Flanagan et al. 1964, Lohnes 1966, Cooley and Lohnes 1968). In discussing the low reliabilities of the speed data, Lohnes (1966) states that these were 'brought about by widespread discrepancies in the timing of the tests in different schools' (pp. 4–9). A similar difficulty was reported in one of the Project Talent follow-up studies (Shaycroft 1967, cited in Cooley and Lohnes 1968), where the poor stability of the speed test scores was attributed to 'anomalies in retest administration' (Cooley and Lohnes 1968, pp. 1–16). Such difficulties arose despite the apparent sophistication of the tests and the careful plans made for their administration (Flanagan et al. 1964). These observations are consistent with what is known about the limitations of group tests when examiners are required to impose several short time-limits in the course of testing (Anastasi 1968). Characteristically speed tests are of short duration (Highsmith 1924, Bernstein 1924, Lord 1956, Flanagan et al. 1964) and are thus particularly prone to unreliability in their administration. The more general limitations of group test procedures have been amply documented (Anastasi 1968, Cronbach 1970, Heim 1970, Vernon 1960) and need not be detailed here. As these aforementioned authors note, they are useful for screening purposes but inadequate for precise measurement. In so far as the measurement of speed is concerned, group tests cannot provide an appropriate basis for measurement.

Group testing, despite its limitations, has been used to provide solution times for individual items by use of special timing devices or other procedures (Sutherland 1934, Slater 1938, Tate 1948, Cane and Horn 1951, Furneaux 1961, Russell 1968). Different techniques have been used to measure these times. Slater (1938), following Sutherland (1934), used three sets of cards numbered 0–9. These cards were placed on a table in such a way that one number from each set was visible to the testees. Subjects were required the record the numbers displayed when an item was completed. One card was turned by hand every 2 s so that a crude item time was measured. Tate (1948) had items individually typed on cards. The subject wrote the answer and the time announced by the testor on each card. Furneaux (1961) employed a mechanical device which was otherwise similar to that used by Slater (1938). Other timing procedures have been used, notably by Russell (1968). In that study, a special cyclometer displayed a set of three numbers, each varying in an apparently random fashion but changing at a fixed rate. At the beginning, and after completing an item, the usbject was required to look at the screen and record the number displayed. By a special decoding procedure, the time to work through a complete item could be computed.

Although such items are reported as 'solution times', this description is far from accurate. Such times represent the duration of a sequence of activities, from turning a page, reading the problem, thinking about its answer, checking the answer, and then recording it, together with any of a number of other irrelevant acts, such as correcting the solution, succumbing to distractions, changing pencils, among others. While obviously better than the gross rate measures commonly used, these times are nevertheless crude. Even using a cyclometer of the type described by Russell (1968), having to read the time adds time which is irrelevant to the problem solving process. Russell (1968) affirms '... it must be recognised that the time score is not a pure measure of the time to solution of individual items'.

Brierley (1969) reports a brief investigation which he conducted into the time of irrelevant activities. In answering the Matrices, he estimates that 'more than 3 minutes may well be spent simply turning pages and writing answers'. It is further assumed that in these procedures the subjects will begin working on the problem immediately it comes into view. Such an assumption has not to this writer's knowledge been supported by appropriate studies.

In an attempt to overcome some of the difficulties introduced by interstitial activities, Furneaux (1961) used a special correction factor. This time constant was individually determined on the basis of subsidiary studies of the same subjects used in the main investigation, the constant being subtracted from the individual item times. While this is a refinement, it still does not ensure accurate individual item times. Indeed, it is not possible to know how Furneaux (1961) determined such a correction factor as he does not give further details. In any case, such irrelevant activity times need to be partialled out of the data for each of the items that is to be used in reaching the answers to the research problems.

Some of the problems inherent in testing large groups can be overcome by testing small groups of four to six subjects. Hunsicker (1925) employed this procedure as did Cane and Horn (1951). However, the item times still included irrelevant components, such as writing the solution and operating the apparatus.

While individual testing overcomes the many problems of group presentation and recording, it does not necessarily remove all of the difficulties. While the tester may be able to adjust the recorded times for some of the interstitial overt acts, it is not possible to correct for covert effects, such as knowing when the subject has actually begun his attempts at finding a solution. The Nufferno Test (Furneaux 1955) procedures require surreptitious timing in an attempt to overcome some of these distorting features. The present author, having used these tests in a clinical setting, is well aware of their timing limitations. Although the stopwatch is concealed, the tester has to record times on a duplicate answer sheet which of necessity, has to remain in the subjects view. Hence, the testee can become aware of the fact that each time a solution is written, the tester records some numbers on another sheet of paper. Inaccuracies also arise when the solutions are presented at a rapid rate. Brierley (1969) has also reported similar experiences. He notes that in addition to subjects becoming aware that they are being timed, it is sometimes difficult to define precisely when the answer is written.

Manual and electrical stopwatches and, more recently, millisecond timers and computer controlled administration are used in individual testing situations. Such techniques introduce problems which may or may not be of consequence. Some of these problems have been highlighted in the previous paragraphs. A further difficulty is the possible differential effect of obvious versus surreptitious timing, or what Furneaux (1955) has called 'stressed' and 'unstressed' speed, respectively. As these procedures have a differential effect on performance (Furneaux 1955), it is necessary to treat the different studies separately: one cannot presume that 'natural speed' has been measured if the fact of timing is obvious.

A number of investigators have used complicated techniques in order to get away from some of the more obvious defects described in the previous paragraphs. Brierley (1969) constructed a special apparatus so that very little time would intervene between successive items. By housing the timing apparatus in a separate room, by arranging for the timing to begin only when the test item was presented, and by enabling the answer to be recorded when an electrical switch was depressed, the time added to problem solving time by the apparatus was trivialized. While such apparatus working time is virtually eliminated, this technique does not of course remove those components of time added by interstitial activities in the subject.

The present writer has employed a computer-based control of item presentation and timing to circumvent the problems of apparatus time. Timing was initiated only when the test item was exposed and terminated when the subject depressed a response button mechanically linked to a microswitch. Inter-item times were of short duration (about .5 s) and constant for all subjects. One major limitation of this approach should also be noted here. The short delay between responding to one item and being confronted with the next means that the item is present even though the subject may not be ready or willing to begin working on it. This therefore introduces a potential interstitial time – one cannot know when the subject actually began working on the item. All that is recorded is response latency. Such a problem could be partly solved by interspersing a message, which in effect tells the subject to press an 'Item Presentation Button' when ready for the next problem. This would not be necessary if the research was concerned with forced presentation or massed practice, although even here it would be difficult to record the time when the subject began working on the problem.

A variety of other procedures is available for the presentation of test material. These take the form of self-contained devices or else apparatus linked to computers (see papers in Elithorn and Jones 1973, Gathercole 1968, Gedye 1966, Miller 1968). However, none of these procedures is specifically designed for gathering accurate solution times.

Figure 1 is an attempt to schematize some of the foregoing discussion. The dashed line at the top represents the hypothesized periods of mental activity concerned with solving the problem. The 'true' problem solving time is the sum of components V, W, X, and Y. 'a' marks the onset of the process, 'b' the point at which a solution is available. The second line represents apparatus time. From the point of view of the procedure, the problem could have been presented anywhere between 'A' and 'B'. Also, the end of timing could take place anywhere between 'C' and 'D', depending on what is required of the subject once he has a solution to offer. For example, if one of a set of keys has to be pushed to record a multiple choice answer, some of the time between 'C' and 'D' will be taken up in locating the correct key.

While the discussion to this point may imply an unrealistic demand for precision, it can be argued that until such time as the measurement problems are minimized, research must be of limited value and any theory based on it inappropriately speculative. It is the impression of this writer that insufficient attention has been paid to the measurement problem and to an analysis of what is supposed to be measured.

The rush into speculation is nowhere stronger than in the factorial analyses of 'speed tests'. Writers have ignored the quality of their data and have proceeded to erect elaborate structures on data which are unworthy of such efforts. The problem is that if mental speed is to be investigated using factorial techniques, then the method of investigation is predetermined by the method of analysis. Proper scientific research requires the opposite. Factorial procedures need large amounts of data. The most economical way to gather such data is by group testing large numbers of subjects. Yet, as

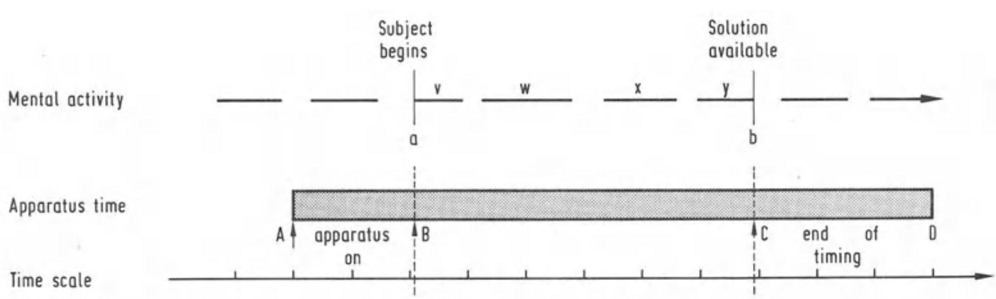


Fig. 1. Hypothesized relationship between problem solving activity and timing of solution

has been argued, group testing cannot, by its very nature, provide the quality of data necessary for the proper investigation of mental speed. This was apparent well before factor analysis achieved its popularity as a data analysis technique in the 1930's. It was made explicit by McFarland in his 1927 review and emphasized by Peak and Boring in their 1926 study.

If there is some substance to the multi-component conception of speed, it is necessary to isolate the speed components for proper measurement. However, because of our current technical limitations, the basic unit of analysis must be the single item time. Such times are subject to distortions, but at least they are minimal. These times should be measured in individual rather than in group test situations as the latter make it almost impossible to cope with distorting factors.

As noted earlier, Furneaux (1961) asserted that item times (response rates), were 'simple and unambiguous' and they could not be easily redefined in terms of simpler determinants. From the foregoing discussion, it is suggested that his assertions are unfounded. Theoretically and practically 'speed' measurement presents a number of important difficulties which are only partly overcome by timing individual items during individual testing. The further assertion that response rates cannot be defined in terms of simpler determinants is also questionable. There is sufficient evidence for instance, that response times are significantly influenced by contextual factors. The problems of speed measurement discussed in this section are basic but by no means exhaustive. The effects of instructions on behaviour during testing and the ways in which such instructions and their various nuances interact with other factors to produce differential outcomes needs to be investigated. Researchers may fail to publish instructions or when reported, no evidence is provided that subjects have responded to instructions in a way that is congruent with the researcher's intentions. Cattell (1971) reports that certain 'speed' factors will emerge only when speed is emphasized in the instructions. Eysenck (1967) has pointed out the need to pay careful attention to structuring the test situation if the effects of neuroticism on performance are to be elicited.

Other factors can be identified as likely to have an effect on measures of speed. The setting in which the research is conducted, the structuring and familiarity or otherwise of the test material, the massing or spacing of the items, and so on need to be considered. These and possibly many other factors serve to complicate speed measurement and if taken into account quickly serve to dispel the simplistic assertions of Spearman (1927), Furneaux (1961), and others. One important aspect not considered so far is item difficulty. This is the subject of the next section.

A Digression into Item Difficulty

For some purposes, the 'difficulty' of an item is not important as individuals can be given the same set of items and their solution times can be compared within items to assess 'mental speed'. However, if 'mental speed' is hypothesized as something general across varied item content, any test of this hypothesis presupposes items of comparable, and hence known, difficulty. This in turn presupposes an adequate bases for difficulty determination.

According to standard procedures, items are considered to be of equivalent difficulty if equal percentages pass or fail the item. While this may be adequate for most of applied measurement, it has one fundamental failing. It does not take account of the possibility that the 50% or so who pass item 'a' may be quite distinct from the 50% who pass item 'b'. Yet 'a' and 'b' would be defined as being of equivalent difficulty. Such an assumption is hard to justify, even if there is some overlap of individuals who pass both items.

Such difficulty indices are of necessity sample dependent; to obtain generalizable difficulty values necessitates representative or substantial random samples. This solution is costly and open to the effects of population structure changes which can invalidate the indices.

Once it is recognized that equating tests for difficulty is problematic, given that the difficulties of the components are crude, subsequent interpretation of test scores becomes questionable. If a battery of tests with varying difficulties is given to a group, then, say, the interpretation of a factor analysis becomes, at the very least, a complicated exercise. If test 'A' is 'easy', then it could involve the use of a different set of skills to those requred for test 'B', a more 'difficult' test. The resulting factor structure might be quite different to the structure which could have emerged with tests of equal difficulty. Findings reported by McDonald (1965) illustrate some of the problems. After carring out a principal components analysis on the test results of two groups of subjects (on the Progressive Matrices), the second component to emerge for one of the groups (the younger of the two) was identified as a difficulty factor. Using a procedure for non-linear factor analysis, McDonald found that the apparent 'difficulty' factor was a curvature component 'not identifiable as such by conventional factor-analytic techniques'. Apart from the questions which these findings raise for factor analysis in general, it appears that the underlying assumption that the items, with some misplacements, increase in difficulty, is not supported, at least not for the younger subjects (mean age 13.96 years).

Furneaux (1961) has suggested that the concept of item difficulty was introduced to account for the introspective observation that the 'sense of effort associated with attempts to solve some problems is stronger than that associated with others'.

In 1903, E.L. Thorndike confronted his contemporaries with a number of problems, the formemost being that of discovering adequate units of mental measurement:

> Educational science needs lists of words in spelling, of examples in arithmetic, algebra and geometry ... etc: so chosen that any one will be of approximately the same difficulty as any other .... The service rendered to physical science by the inch, the ounce, the ohm, the ampere ... should be duplicated in mental science .... Until we have such units all our investigations rest on insecure foundations. (pp. 169–170)

In 1903, Thorndike was optimistic about the emergence of a solution to the difficulty problem, suggesting that 'any trained student' who possessed ingenuity and a 'knowledge of elementary statistics' would overcome the problem of scaling. Years have passed and the problem is still with us (Angoff 1971).

Contemporary test theorists know, as did their predecessors, what they are aiming at. An acceptable scale would be constructed in such a way that if a person passes an item of given difficulty, he will pass all items that are less difficult; if he fails an item of given difficulty, he will also fail any item of greater difficulty.

In his early writings, Thorndike (1903) accepted a measure of relative status as a form of measurement, recognizing at the same time that it was not very satisfactory. As an interim solution, he proposed the use of equal percentage passing as an approximate unit. This proposal was followed up in his later works (Thorndike et al. 1927), but even then, the perfect scale was not achieved. Thus, while the scale then developed (the CAVD) was 'at all points more accurate than the best scales previously available', it still needed to be 'improved by more extensive experimentation' (p. 472).

In developing the CAVD, Thorndike and his colleagues drew a distinction between the item difficulty and intellectual difficulty. In doing so, they diverted attention from the problem of item scaling. Thus they asserted

> ... for every theoretical and practical purpose in the measurement of intellectual difficulty, we should use collections of tasks rather than single small tasks. We ought to measure the difficulty

of single tasks, but we can profitably measure *intellectual difficulty* only in the case of composites which contain enough kinds of tasks to represent a fair sampling of all intellect as it operates at that level .... (p. 133)

Thorndike's reasons for rejecting the single item as the focus of difficulty scaling included their low correlation with his criterion of intellect (the total score), their heterogeneous variance, the restricted range of intellect sampled by such items, and their proneness to being influenced by transient effects and special knowledge. Single items, according to this view, measure 'but a small part of intellect plus a large error' (p. 117). His solution to the difficulty problem was to combine four sets of common content items (10 each from C – sentence completion, A – arithmetic, V – vocabulary, and D – comprehension) to create a composite. Difficulty was then determined by the percentage of his criterion group who passed the composite. For an individual to pass the composite, he had to pass 50% of the items. The difficulty level of the composite could be changed by juggling the items. Thus, a given composite could be made up of items of widely varying 'difficulties'. The Thorndike approach seems somewhat nonsensical in that if a single item contains but a small part intellect and a large part error, a collection of such items may contain a middling part intellect and a substantial part error, unless it can be assumed that none of the error is systematic, an unlikely assumption. Further, given that the items in any composite vary widely in their difficulty measured by percentage passing, two individuals passing the composite can pass by radically different routes. It is possible for individual A to pass the composite by passing 50% of the items whereas individual B might pass by solving the other 50% or all of the items. Nevertheless, a similar procedure was employed for the age level scaling of the Stanford-Binet in its various revisions (Terman and Merrill 1960). Although a different procedure was used for scaling the Wechsler tests, these too, like most tests of intelligence, rest on the common technique of scaling by means of percentage passing.

Subsequent attempts (e.g. Thurstone 1928) to overcome the 'difficulty' problem have met with little success. Gulliksen (1950), in his discussion of item analysis within the framework of classical test theory, has surveyed a variety of procedures for difficulty determination. None has succeeded in overcoming the common limitation of variation in item parameters as a function of the ability level of the group. Such problems have led to a variety of alternatives including the use of multiple indices (e.g. Heim 1970, Anstey 1966), but as yet no substantial advances have been made (Angoff 1971).

A more recent view of the problem of 'difficulty' has been presented in the information theory approach of Newell and Simon (1972). They descirbe their tasks as being 'moderately difficult problems of a symbolic nature' (p. 3). The time taken to solve problems is regarded both as an important aspect of difficulty as well as being an index of difficulty. However, they recognize that 'difficulty' requires reconceptualization in the framework of their approach:

In constructing a theory of problem difficulty we should like to identify those aspects of task environment and the problem solver that are the major determinants of difficulty – whether measured by solution time or any of the alternative measures. (p. 93)

'Difficulty' in the Newell and Simon formulation has to be viewed in terms of the interaction between the task environment and the programme of an information processing system. Task environment is conceptualized as a set of methods for problem solution together with an 'executive structure' for selecting and applying these. An important determinant of difficulty is what Newell and Simon call the problem space. This is the set of possibilities for solving the problem as seen by the problem solver. The methods at the disposal of the solver are used to examine the elements of the problem space one at a time. In the simplest case, the entire problem space is searched using the methods available to the problem solver. In this instance 'time to solution will be roughly proportional to the total size of the space'. The 'problem solving mechanism' proposed by Furneaux (1961) has similar features to that of Newell and Simon. At this point, however, their analysis of problem or item difficulty has no immediate implications for determining item difficulty. The suggestion that it can be roughly indexed by time taken is not novel.

An interesting but somewhat limited approach to item difficulty is presented in the work of Elithorn and his colleagues on the Perceptual Maze Test (Smith et al., interim report, unpublished work; Davies and Davies 1965, Elithorn et al. 1966). The test is made up of a triangular lattice with dots at a number of the intersections. The subject is required to trace a path from apex to base, passing through a given number of dots, while moving in a forward direction.

What is called the 'subjective difficulty' of the maze can be varied in four ways. These are the physical dimensions of the maze, the size of the background lattice, the number of the target dots on the lattice and the arrangement of the dots. Although subjective difficulty can be specified very precisely, the 'difficulty' of each maze appears to be based on the standard 'percentage passing' formula (Smith et al., interim Perceptual Maze Test report, unpublished work).

An alternative approach to assessing the difficulty of each maze was proposed by Davies and Davies (1965). It is based on the idea that each maze has a large number of distinguishable pathways. The 'difficulty' is then related to the number of paths through the maximum number of dots on the solution path. The subject obtains a score based on the 'difficulty' of finding a path through the number of dots attained. This procedure differs from the original in that a graded score is possible for each maze. Elithorn simply scored for pass or fail. Davies and Davies (1965) define 'empirical difficulty' as the percentage of subjects who pass the maze. The two measures 'difficulty' and 'empirical difficulty' correlated +.77 and increased to +.94 if the dot saturation of the lattice and the branches at each choice point on the correct paths were included in the computation.

While the Perceptual Mazes lend themselves well to precise specification of various parameters, the procedures employed are not transferable to the types of test item found in common intelligence tests: they cannot provide a generalized solution to the difficulty problem because item structure in the usual intelligence test is not obviously reducible to the same elements.

Item analysis is concerned with selecting test items in such a way that the test will have certain specified characteristics, in particular, that the final test will have high validities and reliabilities. According to Gulliksen (1950), there are more than twenty methods of item analysis. The determination of item difficulty is a major component in each. Seen from the standpoint of conventional test design, difficulty determination and scaling procedures are important because of their impact on validity and reliability. For most practical purposes in testing, the inadequacies of such procedures have apparently not been crucial, as witnessed by extensive applied testing. However, they have severely restricted the interpretations of test scores and have hampered research on intelligence, again mainly because interpretation is complicated.

Most discussions of 'difficulty' seen to assume that an item or problem has something which can be called 'its difficulty'. Brierley (1969) and others have questioned this:

> The principal reservation one must have concerning the problem of difficulty scaling is that of the reality of an intrinsic item difficulty. (Brierley (1969)

Similar reservations have been expressed, implicitly or explicitly, by several writers (Campbell 1961, 1964; Cane and Horn 1951, Heim 1970). Campbell (1961), for example, has examined the determinants of item difficulty in relation to a number of factors:

1. Extrinsic factors
   a) The context
   b) Familiarity of content
   c) Non-intellectual factors
2. Intrinsic factors
   a) Item qualities (complexity of content, abstractness, novelty)
   b) Item layout

Campbell (1964) also raised the question of the extent to which many items (particularly of the letter series type used by Furneaux) are prone to the effects of chance strategies, which can then have a profound effect on the success of the subject. Taking series items as one example, she points out that varied solution rules are available but only one is, in the view of the test constructor, correct. Brierley (1969) cites an additional factor, that of irrelevant information introduced to complicate the item. As an instance, he suggests a letter series item in which the rule is given by alternate letters, the others merely serving as distractors. Cane and Horn (1951) found that the position of an item in a test did not affect its conventional difficulty. However, the time spent on the item was related to its position. Open-ended questions led to shorter solution times than did multiple-choice items. Dunn and Goldstein (1959) varied several item features (number multiple choices per item, irrelevant cues, grammatical changes), and found that while they produced changes in conventional difficulty, validity and reliability of the items were not appreciably changed.

A number of studies by Heim and her colleagues (see Heim 1970) have examined a variety of influences on conventional difficulty as well as difficulty indexed by solution speed. For example, Heim (1955, 1957) has found that successful sulutions are proportional to the difficulty of the context: an easy test preceding a more difficult test led to a smaller percentage correct in the latter. Speed was also affected in that easy items would be answered more slowly if they were preceded by a hard test. However, the less intelligent subjects produced inconsistent findings.

A further issue is the extent to which 'complexity' and 'difficulty' are synonymous. Porebski (1954) suggests that they are the same but that complexity may be introduced by irrelevant features or by requiring the subject to have at his disposal other skills (e.g. a 'good' short-term memory). The information theory analysis of Newell and Simon (1972) presents a different conceptual analysis of the same issues.

These various assessments of the effects of contextual factors on conventional difficulty indices provide some support for the view that such indices are sensitive to context. However, they do not resolve the issue as to the independent existence of 'intrinsic difficulty'. Were it not for the work of Furneaux (1961), Elithorn and his colleagues (Smith et al., unpublished work), and Davies and Davies (1965), it would be possible to reject the notion of intrinsic difficulty and simply focus on conventional indices and their empirical relationships and determinants.

The work of Elithorn, and Davies and Davies does, however, point to three classes of difficulty, one of which appears to be 'intrinsic difficulty'. Smith et al. (unpublished work), distinguish between subjective and empirical difficulty. As was noted previously, the subjective difficulty (complexity?) can be varied along any or all of four dimensions (physical dimensions, lattice size, number of dots, and arrangement of dots). Empirical difficulty refers to conventional indices. The Davies and Davies (1965) procedure for indexing the properties of an individual maze leads to what they call calculated difficulty. This index completely describes the structural properties of a given maze. The important point is that the calculated difficulty is invariant. It depends entirely on the maze itself and not on the context or on any other factor extrinsic to the item. In this sense, it is possible to view each maze as having an intrinsic property which appears to be directly related to what is generally regarded as 'difficulty'. It is unfortunate that Davies and Davies (1965) did not attempt to examine the scalability of the calculated difficulties but simply treated them as ranks. Also, as noted earlier, conventional items do not readily lend themselves to such structural specification although it might yet be possible to treat, say, letter series as number of alternative correct solutions and the number of elements in the items as well as the distractors. It might be possible to examine the predictive power of such an index, the goal being to demonstrate that such indices conform to the ideals of scaling described by Angoff (1971).

Furneaux's (1961) attempted solution of the difficulty problem is important because it circumvents the major limitations of conventional difficulty scaling. His solution will be considered later. At this point it is sufficient to note that his procedures are claimed to produce indices which are independent of the standardization group, that they are based on an unambiguous measure, solution time, and that his indices enable the prediction of performance at different levels of difficulty.

Dissatisfaction with classical approaches to mental test scoring has led to the development of a number of different models (Lord and Novick 1968, White 1973a, b; Van der Ven 1971, 1974; Iseler, 1970, Lord 1974, among others). The major shift of emphasis in all these more recent approaches is away from the older deterministic models to approaches based on probability (Lord and Novick 1968, Lord 1974). Of the newer models, those of White (1973a, b; see Chap. 3), Iseler (1970), and Van der Ven (1974) have been particularly influenced by Furneaux's conceptual analysis, but have not attempted to apply his scaling procedures.

Possibly the most widely investigated of the current approaches are those subsumed under the generic title of 'item characteristic curve theory' (icc) (Lord 1974). These models usually have two basic components, the ability of the individual and a vector containing parameters that fully characterize the item. It is then assumed that the probability of a correct response depends only on the level of ability of the individual and on the item parameters. In its simplest form, no assumptions are made about any characteristics of the individual. More complex forms of the model have been developed by a number of writers (e.g. Birnbaum 1968 in Lord and Novick 1968). For example the 'three parameter' version of the model has as its parameters the discriminating power of the item, item difficulty, and the probability of a correct answer for individuals at the lowest levels of ability.

The development of these models, has, according to Lord (1974), been held back at the empirical level because of the problem of estimating the characteristic curve of individual items. These problems appear now to have been overcome and there is accumulating evidence on the validity and usefulness of this approach. However, it appears that more research is still needed.

Mental Speed

Problems in the determination of item difficulty and the conceptualization and measurement of speed greatly limit the value of much psychological work on speed. Factor analytic studies in particular are of marginal interest because of their cavalier disregard of these problems. There are, however, a number of approaches to the study of intelligence which, to a greater or lesser extent, have attempted to cope with these problems and which, at the same time, manifest a desire to conform to conventional notions of scientific theory and research. These are considered next.

Thurstone, in a theoretical paper published in 1937, examined the relationship between ability, motivation, and speed, with a view to appraising 'ability as power' independent of speed and motivation. As part of his model, he proposed a three-dimensional surface with difficulty defining one axis, and response time and probability of success defining the other axes.

One of the major problems, according to Thurstone, was that associated with conventional definitions of speed as the number of easy tasks completed in unit time. As he saw it, the problem was whether high speed can be taken as an index of the ability of complete more difficult tasks without there being any time limits. The three dimensional surface (see Fig. 2) is generated by assuming that a subject with fixed motivation attempts a large number of tasks at each level of difficulty. All tasks are of standard type. Their difficulty is calibrated on some scale using the percentage correct in a standard group as the index of difficulty. From Fig. 2 it can be seen that for a fixed response time, any increase in difficulty will lead to a decrease in proportion correct; as the amount of time allowed is increased, with difficulty held constant, the probability of success will increase. It will also be seen that, like Furneaux (1961), normal ogives are assumed to express relationships between response time and probability of success, and between difficulty and probability of success.

The plane AB in Fig. 2 is at a point corresponding to a large valve of T, response time. The ability surface (curve) at this point has a median indicated by point C and this in turn corresponds to a difficulty level at which the probability of success is .5. According to Thurstone, if T is already a generous time allowance, any further incrase in T will have relatively little effect on the 'psychometric curve' ACB. On the basis of these assumptions, Thurstone is then able to propose a definition of 'power or altitude which is independent of the speed of any performance'.

> The ability of an individual subject to perform a specified kind of task is the difficulty E at which the probability is that he will do the task in infinite time. [Original in italics]

A practical procedure for determining E by means of interpolation is described by Thurstone. That is, various points can be determined experimentally, for example, by measuring response times at different difficulty values.

Brierley (1969) has criticized this approach on several grounds. For instance, the model disregards the effects of continuance at high levels of difficulty, which is likely to produce distortions in the data. These distortions would in turn affect the adequacy of any interpolation which is attempted. As a consequence, the ability surface becomes much more complex than Thurstone realized. A further limitation of Thurstone's model is the failure to be consistent in his definition of *P*. At one point he refers to *P* as the probability 'that the individual subject will successfully complete

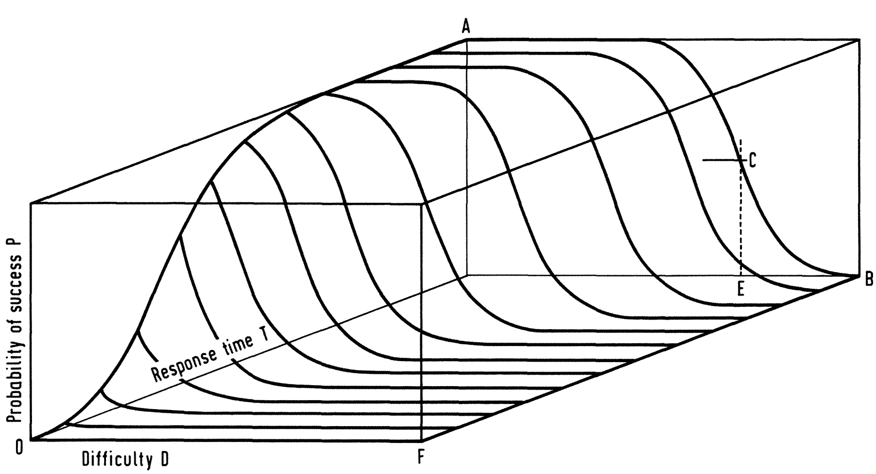


Fig. 2. Thurstone's (1937) 'Ability Surface'

a task' and later, he refers to P as 'the proportion of successful solutions'.

Despite its limitations, Thurstone's model, with the exception of 'power', can fit Furneaux's theory quite closely (Brierley 1969). It gives rise to three dimensions which can be scaled (accuracy, time, difficulty) and which enable an ability surface to be generated. Any given point on the surface will depend on the time limits (defined externally or internally), item difficulty, persistence, and so on. However, it seems highly unlikely that such a surface would ever be achieved in practice for any individual. It would be very difficult indeed to so manipulate practical testing so as to achieve error-free solutions at varied item difficulities, for example.

Brierley (1969) argues for a multidimensional concept of power based on time, difficulty, and accuracy. He suggests that if reliable and practicable units can be found for these dimensions, the generation of an ability surface is possible. Drawing on the power concept in physics, Brierley defines 'power' as the work done in unit time. It is equal to the product of the number of unit solutions achieved and difficulty $\times$ time^{-1}. This formulation does have a number of practical obstacles. The most important of these is that of defining a unit task. Although there has been an attempt to fractionate test items into units (Restle and Davis 1962), its outcome was suggestive rather than definite. Brierley therefore accepted that for the present, it will have to be assumed that items of approximately equivalent difficulty will have to be used. The second major problem, that of difficulty scaling and the determination of item difficulty, might be resolved by using Furneaux's (1961) procedures. Using this model of power, Brierley (1969) was able to demonstrate that diagnosed neurotics performed significantly less efficiently than did normal subjects. The work of both Thurstone and Brierley would suggest that 'power' is a complex concept without a generally accepted definition (see Heim 1970).

Cattell (1971) has attempted to draw together research on human abilities and his views on speed and its relationship to ability and personality will be discussed here. Reference will also be made to two major reviews published by Horn (1970, 1972), who has been particularly important in developing Cattell's theoretical and empirical approach.

Cattell provides a list of, at this stage, tentative empirically based primary abilities. Among those factors which receive the strongest confirmation and which also have comparatively substantial variances are:

UI(4) Perceptual speed (identified in more than than 30 studies) based on tests involving the comparison of similarity in visual material and configuration, mirror reading, and dial recognition.

UI(5) Speed of closure (visual cognition, gestalt perception) based on nine studies including such tests as street gestalt and speed of dark adaptation.

Among the 'lesser, narrower, less substantiated primaries' is

UI 71 Motor speed

At least five broad factors emerged when the intercorrelations among primary factors were further investigated in the ability realm. Cattell (1971, p. 106) identified these as

Fluid general intelligence
Crystallized general intelligence
Power of visualization
Retrieval capacity of general fluency
Cognitive speed

Associated with this pattern is the theory of crystallized (g_c) and fluid (g_f) abilities, which asserts that there are two major attributes which have a significant impact on performance on intellectual tasks. These influences operate somewhat independently (Horn 1972) throughout development and are said to represent the basic components of intelligence. Fluid intelligence is manifested primarily in tasks which are relatively uninfluenced by culture. They are either novel or overlearned. Crystallized intelligence reflects the individual's use of concepts and aids derived from the culture. Both involve the same processes of reasoning, relation perceiving, abstracting, and the like but differ mainly in the extent to which they involve culture specific learning. Some differences exist between Horn and Cattell (Horn 1972). In Cattell's views, g_f is regarded as more explicitly related to hereditary-physiological influences. g_f and g_c have been studied in relation to a variety of associated functions, the details of which are presented in the reviews by Cattell and Horn. For example, there is evidence that they show a differential age decline as well as being differentially influenced by damage to the nervous system. In terms of actual tests, the Matrices load on g_f whereas verbal tests such as the Mill Hill Vocabulary have been found to load on g_c. Cattell (1971 p. 107) notes that the Furneaux Speed and Level Tests are most strongly loaded on g_f.

The third of the major factors, visualization covers a variety of performances involving spatial manipulation (e.g. form boards) and includes 'flexibility of closure' and 'perceptual speed' as correlates.

The fourth broad factor involves retrieval of material 'from memory storage', sometimes called fluency but not analogous to the primary factors of the type which are loaded by word fluency tests.

The fifth broad factor, speed of cognitive (g_s) performance, poses the greatest difficulty in trying to characterize it. As Cattell points out, it has a long history, and is the factor supposedly discussed by Spearman. Its 'existence' was detected in the 'early and thorough' (Cattell 1971, p. 107) studies of Bernstein (1924). Parenthetically, it should be noted that it is extremely hazardous to attempt to place any reliance on Bernstein's study, a hazard which Cattell ignores. The quality of Bernstein's research is such as to make any conclusions derived from it exceedingly suspect.

Cattell notes that an early conception of g_s placed it in the realm of personality-temperament factors and that recently, Horn has suggested that it is an index of motivational strength operating in the actual test situation. Cattell's views are somewhat complex.

The correlational evidence according to Cattell (1971 p. 64) points to a number of different speeds rather than a single general speed. In addition, our conception of speed is complicated by both semantic confusion and the variety of scoring procedures used. The two major indices, number correct and time taken, generally correlate positively. More substantial correlations are found between scores on timed and untimed tests provided that subjects are asked to work quickly and that they are homogeneous with respect to age (Cattell 1971, p. 65). However, under speeded conditions, when intelligence and the effects of other primary factors are partialled out '*two or three generalized speed measures remain*'. Even when such corrections are not made and the speed stress is not introduced, a tempo factor still emerges represented by the work of Rimoldi (1951) and associated with the personality factors UI 30 (aloof independence) and UI 33 (depression-elation).

In addition to tempo, Cattell implicates two further sources of speed difference. The first is identified by UI 22 (corteria), the characteristic level of cortical alertness of the individual, and the second UI 16 (assertive ego), is manifest as ambition in the test situation.

In summarizing the evidence, Cattell asserts that anything that is *general* in cognitive speed [the cognitive speed factor of Bernstein (1924)] is temperamental or motivational in origin and is associated with UI 22 or UI 16. However, the UI 22 temperamental component 'actually extends in a confusing fashion along the frontier between ability and temperament traits' (p. 65). Even though its contribution to 'variance in *high level* abilities is quite small', Cattell suggests that it should be included in any discussion of abilities in that it has an influence on perception and executive performance. In attempting to clarify his distinction, Cattell thus posits temperamental-cognitive speed as distinct from the speed which arises through temperamental tempo, motivation level, and mood. Its peculiar property is that it 'appears only when ability scores are made under "speed" instructions and in scoring a timed performance' (p. 65). In a later discussion of this component, Cattell maintains that it affects speed in a broad spectrum of abilities, such as numerical performance, social skills, perceptual speed, and especially mechanical speeds such as writing (p. 107). This component in turn makes only a minor contribution to speed in intelligence-demanding tasks, or what Cattell refers to as 'power intelligence'. This further type of speed is largely an expression of the same ability as is measured in fitness and error-freeness of response'. This conception of speed is similar to Spearman's view and is supported, according to Cattell, by it being located within g_f rather than g_s in the second order factor pattern. It is the component identified by Furneaux as intellectual speed. The nature of this conception of speed is best illustrated by the following:

> ... By any reasonable perspective this simple speed factor is a distinctly broader factor even in the cognitive realm itself, than are the two intelligences. For example it operates even more obviously in mechanical and perceptual performances than in intelligence. Speed measured in successful, intelligence problem-solving is local to intelligence (being zero if a person cannot solve the problem!). If intelligence is considered speed at all, it is speed in more complex performances than those that are typically strongly loaded by g_s. (Cattell 1971, p. 108)

Three important features emerge from this brief examination of Cattell's work. Firstly, his approach is essentially structural in that it attempts to isolate the major elements and examine their interrelationships. Secondly, his review of relationships among the ability factors points to the possibility of two conceptions of speed, that contained in g_f and that in g_s at the second order. (It is unfortunate that Cattell has called g_s 'cognitive speed' as this factor is loaded by tasks such as 'writing speed' and 'cancellation speed'). Finally, Cattell attempts to include his personality factors in discussing the relationships.

The decades which followed the establishment of intelligence as an important psychological concept have, with few exceptions, witnessed major and minor controversies about one or other of its facets. While it is unlikely that many of the problems will be answered in the foreseeable future, some of the more fundamental issues have at least been identified. Butcher (1968) has pointed out that from the point of view of the scientific study of intelligence, what is needed is a law or set of laws which can act as the basis for major advances in our understanding of intelligence. Such laws would, according to Butcher, help establish an acceptable definition of intelligence which would then facilitate further developments. The work of Furneaux (1961) was seen as a potentially important approach in relation to this problem.

The problems inherent in conventional tests led Furneaux (1961) to conclude that some other approach to intelligence testing should be devised. In effect, this involved 'setting on one side the whole of the approach to cognitive function which originated with Binet' and which had 'come to be taken for granted ever since'. The alternative approach devised by Furneaux was based on evidence (Slater 1938, Furneaux 1948, Tate 1950) that studies of response rate were 'simple, unambiguous and theoretically and practically relevant'. Further, measures of response rate also appeared to be such that they were not easily 'redefined in terms of sets of simpler determinants ....' These somewhat bold assertions were considered, and found wanting, in an earlier section of this chapter. Nevertheless, Furneaux was able to use this analysis of speed as part of a seemingly successful approach to the determination of item difficulty. Indeed, Furneaux's difficulty scaling procedures not only take account of some of the criticisms made of the total score unit, but also show that difficulty and log time to correct solution are linearly related. The import of this relationship is expressed as follows:

> The increase of 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. (Eysenck 1967)

Despite the apparent significance of this finding, there were no attempts at replication until the late 1960's, even though Furneaux had begun to publicize his work as early as 1948 (Furneaux 1948, 1950, 1952). This delay was partly due to the length of time it took for the actual procedures to be published (Furneaux 1961) and because, as Butcher (1968) remarked, the paper which reported this main finding was 'decidedly obscure' in its presentation.

Furneaux treats problem solving (item solution) as a special case of multiple-choice reaction time. The problem solving process requires, as one of its components, a search for the solution. As conceptualized by Furneaux, the search process involves a series of operations each component of which occupies a fixed time. The more difficult the problem, the greater will be the number of components required and thus an increase in the time required for the search. Furneaux developed this model on the basis of work described by Hick (1952), Hyman (1953), and others. These studies had demonstrated that choice reaction time (RT) was linearly related to the complexity of the choice situation. This complexity can be transformed into an information-theory unit, the bit. In the case of simple reaction time, 0 bits of information are involved. One bit of information is present when two alternatives are presented, 2 bits when there are four possibilities, 3 bits when there are eight choices, and so on. By using a choice RT task involving 0–2 or more bits, it is possible to compute a 'rate of gain' measure which is the slope of the best fit line through the mean RT for each set of choices.

While simple RT does not show a strong relationship with intelligence, there is evidence of a relationship between intelligence and the rate of gain measure in choice RT (Roth 1964). This evidence is seen by Eysenck (1967a) as indicating that reaction time experiments do not, as previous studies suggested, contradict a theory in which speed is a central concept. Also, Roth's research has important implications for the choice reaction time model of problem solving put forward by Furneaux. (See chap. 4 by Jensen.)

Furneaux's work – which was stimulated to some extent by Eysenck, who in turn was influenced by the subsequent development of Furneaux's ideas – displays a number of important characteristics. Firstly, it presents, albeit in an intricate and at times obscure manner, a 'solution' to the computation of item difficulties. Secondly, it develops an operationalized conceptual scheme which differentiates aspects of problem solving that in the past were either inexact or unremarked, and then proceeds to examine empirically the relationships among these. Next, it incorporates speed as a central concept. Fourthly, it is firmly lodged within the conventional scientific framework of psychology and, finally, it makes provision for linking problem solving with personality variables. Some of its main strengths, however, are the main source of its major deficiencies. The complexity and occasional obscurity of the difficulty scaling procedures for individual items, the keystone for empirical testing, make independent replication exceedingly cumbersome and even with the aid of computers, very time-consuming. The fact that Furneaux was able to use his proceudres to compute difficulty indices for his research test items using only an electromechanical calculator is a remarkable instance of persistance, or as he would have it, continuance. Furneaux's approach is detailed in a chapter written in 1961 and reprinted in Eysenck (1973), so that no attempt will be made to present it here.

Brierley (1969) and the present writer (Berger 1976) have independently attempted to follow Furneaux's prescriptions for difficulty scaling and both were only partially successful. A crucial test of his method centres on the establishment of a linear relationship between solution time and difficulty. Further, for subgroups of subjects defined empirically on the basis of their scores on an index of 'speed', the slope coefficients of the linear function should not differ significantly. It is somewhat surprising that despite adequate statistical testing of most of his other hypotheses, these particular aspects of Furneaux's work were not assessed statistically, either in terms of slope or of departure from linearity. Further, on close inspection of his procedures, it emerges that for some of the items, difficulty estimates have to be established subjectively, making it possible to select values which assist in producing a linear-looking trend in the data. A further limitation to Furneaux's work is that it is cumbersome, even with the aid of a computer. At each stage a variety of decisions has to be made and a new set of estimates submitted to trial analysis. These scaling procedures are, to say the least, laborious. All these limitations when combined make the use of Furneaux's scaling procedures less attractive than when first encountered.

Despite these non-trivial limitations, the present writer (Berger 1976) was able to partially replicate Furneaux's findings of a 'linear-looking' relationship between speed and difficulty for subsets of items from the Mill-Hill and Advanced Progressive Matrices Tests. It will therefore be interesting to see if different models of problem solving are able to further support Furneaux's conclusions and Eysenck's assertions about speed-difficulty relationships. There is no doubt, however, that Furneaux's work is an important contribution to the analysis of test performance and conceptually at least it provides a useful basis for the further development of theory and research in problem solving. It is also worth noting here that psychologically the model is limited in that it fails to take into account the dynamic nature of problem solving, such as the likelihood that people learn about solving test items in the process of being tested; their characteristics as problem solvers change as testing progresses.

A number of aspects of Eysenck's approach have already been noted. In essence, it consists of two major themes. Firstly, he has focused on a careful analysis of the deficiencies of conventional approaches to measurement. These have already been considered. The second theme is an attempt to construct a model of intelligence test performance which, influenced in part by Furneaux, gives speed a central role. There are as well a number of other important characteristics. The underlying aim is to establish laws of behaviour. The approach emphasizes the relationship between personality and test performance and several investigators have sought such relationships. Like much of Eysenck's other work, there is an attempt to develop the theory on several levels, via information theory, psychophysiological measurement (evoked potentials), and genetic analysis. All the above-mentioned features are observable in his editorial comments to a compendium of papers published in 1973 (Eysenck 1973) and will not be detailed here. Instead, attention will be given to a number of issues which support or challenge his views.

Brierley (1969), with some justification, comments on the 'static' nature of Eysenck's (1967) 'cube' model of intelligence, initially described in 1953 and 1967. Although serving a different purpose to that of Guilford's structure of intellect cube, it is equally lifeless, and at variance with the complexities of human problem solving, seen for instance in the work of Newell and Simon (1972). Eysenck's reliance on Furneaux's claims, given the questionable empirical status of the latter, appears somewhat optimistic. Further, his incorporation of the work of Roth (1964) was possibly premature because of the conceptual and methodological limitations of Roth's work (Berger 1976). (It will be recalled that Roth found a statistically significant correlation between a measure of 'rate of gain' of information derived from choice RT and one of intelligence. See Chap. 4 by Jensen.)

The present writer also attempted to replicate Roth's findings (Berger 1976). Although the resulting data are also derived from a study with some methodological limitations, it became obvious that any attempt to assert generally that there is relationship between 'rate of gain of information' and intelligence is simplistic. Directionally, correlations between a rate of gain index and IQs on the Mill Hill and Matrices were consistent with Roth's findings. On one-tailed tests, however, only the Mill Hill (vocabulary) was found to be significantly correlated. However, the shared variance amounted to no more than 3.2% in a sample of over 100 subjects with age partialled out. It is possible that much of Roth's correlation was carried by the verbal component of his test battery. It was also found that the correlations between rate measures and Furneaux indices of speed on the Mill Hill and Matrices were not significant, despite the sample size. These results, as well as theoretical considerations, suggest the presence of some complex relationships, and the absence of others, which would require a theory differentiated to a much greater extent than is the case with the proposed by Eysenck.

One of the strengths of Eysenck's approach is his ongoing attempt to link personality variables, particularly extraversion (E) and emotionality (N), to test performance. The present writer's work on the Furneaux approach included measures of E and N and thus enabled the testing of a number of predictions derived from Eysenck's theory of personality-test performance relationships. Before summarizing the findings, it is important to note that the theory requires a number of conditions to be satisfied before hypothesis testing is regarded as valid. For instance, the observed impact off E or N is likely to be a function of the degree of each in the sample. An ambivert group and crudely measured variables would not be regarded as an adequate combination with which to test hypotheses. As the main thrust of the wirter's study was not directed specifically at testing personality-test performance relationships, the degree of support (or lack of support) found should not be regarded as strong evidence.

In support of Eysenck's predictions, it was found that E and Mill Hill scores were negatively correlated ($r = -.36$, one-tail test for N's greater than 100 – McNemar 1969, $P < .001$). This correlation is linear and was not affected by age in the sample of male subjects. No relationship was found between E and Matrices, and none of the time-based measures, including Furneaux speed scores, were correlated with E. For the Mill Hill, increasing E was significantly associated with increasing errors ($r = .24$) and abandonments ($r = .23$). For the N dimension, several significant correlations with the study variables disappeared when age was partialled out. In summary, the findings of the study were such as to justify more detailed and specifically tailored research on Eysenck's views of personality-test performance relationships.[1]

Concluding Remarks

The problem of speed remains – in Cattell's (1971) words, a 'vexed question'. It is destined to remain so until the conceptual, theoretical, and operational difficulties are overcome. Research on the latency of evoked potentials (see Chap. 6 by Hendrickson), in conjunction with the work of White (see Chap. 3), seems to be a potentially fruitful combination, likely to be complemented by linking it to the framework of a theory of personality such as that being evolved by Eysenck (1981).

In retrospect, the major contributions of Furneaux are possibly his attempts to achieve conceptual refinement and his examination of the implications of his ap-

1 Also relevant, inter alia, are the studies by Beauvale (1977), Eysenck and White (1964), Gibson (1975), Goh and Farley (1977), Klein et al. (1976), Krol (1977), Lienert (1963), Mohan and Kumar (1976), Walsh and Walsh (1978). (Editorial Footnote.)

proach for the study of test performance. The insights he provides make one wonder at the cavalier way in which psychologists have subjected their data to intricate analyses which they then, for instance, assume tells them something meaningful about the structure or heritability of intelligence. The unquestioned assumption that a mass of data can compensate for poor quality, lack of conceptual clarity, and technical inadequacy is even more remarkable.

In retrospect too, the legacy of the small group of researchers who espoused a 'scientific approach to intelligence' is to be found in their search for clarity in conceptualization and for data which were of sufficient quality and relevance to show that they respected their ideas. What they thought and did may today seem somewhat crude, given our more complex computer-inspired models and millisecond timing devices, but they did nevertheless provide a base for later developments. One has to recognize as well that the framework within which they functioned, namely, their particular understanding of science, was narrow and stultifying. Our views of what constitutes science continue to change and are becoming less constricting (Chalmers 1978). Hopefully this will have a liberating effect on our attempts to understand the structures, processes, and experiences which give rise to individual differences in human performance.

References

Anastasi A (1968) Psychological testing, 3rd edn. MacMillan, London

Angoff WH (1971) Scales, norms and equivalent scores. In: Thorndike RL (ed) Educational Measurement, 2nd edn. American Council on Education, Washington

Anstey E (1966) Psychological tests. Nelson & Sons, London

Beauvale A (1977) Effect of intelligence on relations between temperamental variables and semiprogrammed learning performance. Pol Psychol Bull 8:83–89

Berger M (1976) A study of the Eysenck-Furneaux approach to the analysis of performance on intelligence tests. PhD thesis, University of London

Bernstein E (1924) Quickness and intelligence, an inquiry concerning the existence of a general speed factor. Br J Psychol [Monogr]

Bills AG (1931) Blocking: a new principle of mental fatigue. Am J Psychol 43:230–245

Birren JE (1964) The Psychology of Aging. Prentice-Hall, Englewood Cliffs

Boring EG (1957) A History of experimental psychology. 2nd edn. Appleton-Century-Crofts, New York

Botwinick J (1973) Aging and behavior. Springer, Berlin Heidelberg New York

Brierley H (1960) The speed and accuracy characteristcs of neurotics. MA thesis, University of Liverpool

Brierley H (1969) The intellectual speed and accuracy of psychiatric patients diagnosed as neurotic. PhD thesis, University of Leeds

Burt C (1909) Experimental tests of general intelligence. Br J Psychol 3:94–117

Burt C (1949) The two-factor theory. Br J Psychol II:151–179

Butcher HJ (1968) Human intelligence. Its nature and Assessment. Methuen, London

Campbell AC (1961) Some determinants of the difficulty of non-verbal classification items. Educ Psychol Meas 21:899–913

Campbell AC (1964) Intelligence test items and serendipity: further considerations. Aust J Psychol 16:102–106

Cane VR, Horn V (1951) The timing of responses to spatial perception questions. QJ Exp Psychol 3:133–147

Cattell RB (1971) Abilities: their structure, growth, and action. Houghton Mifflin, Boston

Chalmers AF (1978) What is this thing called science? Open University Press, Milton Keynes

Cooley WW, Lohnes PR (1968) Project Talent: Predicting Development of Young Adults. American Institutes for Research and School of Education, University of Pittsburgh, Pennsylvania, United States

Cronbach LJ (1970) Essentials of psychological testing, 3rd; Internation edn. Harper & Row, New York

Davies ADM, Davies MG (1965) The difficulty and graded scoring of Elithorn's perceptual maze test. Br J Psychol 56:295–302

Du Bois PH (1970) A history of psychological testing. Allyn and Bacon, Boston

Dunn TF, Goldstein LG (1959) Test difficulty, validity and reliability as functions of selected multiple choice construction principles. Educ Psychol Meas 19:171–179

Elithorn A, Jagoe JR, Lee DN (1966) Simulation of a perceptual problem-solving skill. Nature 211:1029–1031

Elithorn A, Jones D (1973) Artificial and Human Thinking. Elsevier, Amsterdam

Eysenck HJ (1939) Primary mental abilities. Brit J Educ Psychol 9:270–275

Eysenck HJ (1967) Intelligence assessment: a theoretical and experimental approach. Br J Educ Psychol 37:81–98
Eysenck HJ (ed) (1973) The measurement of intelligence. MTP Press, Lancaster
Eysenck HJ (ed) (1981) A model for personality. Springer, Berlin Heidelberg New York
Eysenck HJ, White PO (1964) Personality and the measurement of intelligence. Br J Educ Psychol 34:197–202
Farnsworth PR, Seashore RH, Tinker MA (1927) Speed in simple and serial action as related to performance in certain intelligence tests. Ped Sem 34:537–551
Flanagan JC, Davis FB, Dailey JT et al (1964) The American high school student. University of Pittsburgh Press, Pittsburgh
Freeman FN (1939) Mental tests: their history, principles and applications. Harrap, New York
Firth CD (1973) Nature of rest pauses in a simple tapping task. Percept Mot Skills 36:437–438
Furneaux WD (1948) The structure of "g" with particular reference to speed and power. In: Proc 12th Int Congr Psychol.
Furneaux WD (1950) Speed and power in mental functioning. Paper to BPS, London.
Furneaux WD (1952) Some speed, error, and difficulty relationships within a problem solving situation. Nature 170:37
Furneaux WD (1955) The Nufferno tests of speed, accuracy and level. University of London Press, London
Furneaux WD (1961) Intellectual abilities and problem solving behaviour. In: Eysenck HJ (ed) Handbook of abnormal psychology. Basic Books, New York
Gathercole CE (1968) Automation techniques in clinical psychology. BPS Bull 21:161–165
Gedye JL (1966) Experience with an automated psychological system, designed for clinical use. Paper read at BPS Annu Conf
Gibson HB (1975) Relations between performance on the advanced matrices and the EPI in high-intelligence subjects. Br J Soc Clin Psychol 14:363–369
Goh DS, Farley FH (1977) Personality effects on cognitive test performance. J Psychol 96:111–112
Goodenough FL (1935) The development of the reactive process from early childhood to maturity. J Exp Psychol 18:431–450
Guilford JP (1967) The nature of human intelligence. McGraw-Hill, New York
Guilford JP, Hoepfner R (1971) The analysis of intelligence. McGraw-Hill, New York
Gulliksen H (1950) Theory of mental tests. Wiley, New York
Harré R, Secord PF (1972) The explanation of social behaviour. Blackwell, Oxford
Hartshorne H, May MA (1928) Studies in the nature of character, vol 1. Macmillan, New York
Heim AW (1955) Adaptation to level of difficulty in intelligence testing. Br J Psychol 46:211–224
Heim AW (1957) Psychological adaptation as a response to variations in difficulty and intensity. J Genet Psychol 56:193–211
Heim AW (1970) The appraisal of intelligence. NFER, London
Hick WE (1952) On the rate of gain of information. QJ Exp Psychol 4:11–26
Highsmith JA (1924) Relation of rate of response to intelligence. Psychol Monogr 34:1–33
Horn JL (1970) Organisation of data on life-span development of human abilities. In: Goulet LR, Baltes PB (eds) Life-span developmental psychology. Academic Press, New York, pp 182–198
Horn JL (1972) The structure of intellect: primary abilities. In: Dreger RM (ed) Multivariate personality research. Claitor, Baton Rouge
Hunsicker LM (1925) A study of the relationship between rate and ability. Columbia Univ Contrib Educ:185
Hyman R (1953) Stimulus information as a determinant of reaction time. J Exp Psychol 45:188–196
Iseler A (1970) Leistungsgeschwindigkeit und Leistungsgüte. J Betz, Berlin
Jones HE (1959) Intelligence and problem solving. In: Birren JE (ed) Handbook of aging and the individual. University of Chicago Press, Chicago
Klein GA, Blockovich RN, Buchalter PS, Huyghe L (1976) Relationship between reflection-impulsivity and problem solving. Percept Mot Skills 42:67–73
Krol TZ (1977) The relation between nervous system strength and aspiration level. Pol Psychol Bull 8:99–105
Laming DRJ (1968) Information theory of choice reaction time. Academic Press, London
Lawley DN, Maxwell AE (1963) Factor analysis as a statistical method. Butterworth, London
Lemmon VW (1928) The relation of reaction time to measures of intelligence, memory and learning. Arch Psychol 94
Lienert GA (1963) Die Faktorenstruktur der Intelligenz als Funktion des Neurotizismus. Z Exp Angew Psychol 10:140–159
Linden KW, Linden JD (1968) Modern mental measurement: a historical perspective. Houghton Mifflin, Boston
Lohnes PR (1966) Project talent: measuring adolescent personality. School of Education, University of Pittsburgh, Pittsburgh
Lord FM (1956) A study of speed factors in tests and academic grades. Psychometrika 21:31–50
Lord FM (1974) Individualized testing and item characteristics curve theory. In: Krantz DH, Atkinson RC, Luce RD, Suppes P (eds) Measurement, psychophysics, and neural information processing. Freeman, San Francisco

Lord FM, Novick MR (1968) Statistical theories of mental test scores. Addison-Wesley, Reading MA

Mangan GL (1959) A factorial study of speed, power and related temperament variables. Br J Educ Psychol 29:144–154

McDonald RC (1965) Difficulty factors and nonlinear factor analysis. Br J Math Stat Psychol 18:11–23

McFarland RA (1928) The role of speed in mental ability. Psychol Bull 25:595–612

McNemar Q (1969) Psychological statistics. Wiley, New York

Miller E (1968) A case for automated clinical testing. BPS Bull. 20:7–10

Mohan V, Kumar D (1976) Qualitative analysis of the performance of introverts and extraverts on standard progressive matrices. Br J Psychol 3:391–397

Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs NJ

Peak H, Boring EG (1926) The factor of speed in intelligence. J Exp Psychol 9:71–94

Porebski O (1954) A psychological and statistical study of speed and power as variables of human ability. Occup psychol 28:218–231

Restle F, Davis JH (1962) Success and speed of problem solving by individuals and groups. Psychol Rev 69:520–536

Rimoldi HJA (1951) Personal tempo. J Abnorm Soc Psychol 46:283–303

Roth E (1964) Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz, Z Exp Angew Psychol 11:616–622

Ruch GM, Koerth W (1923) 'Power' vs. 'speed' in Army Alpha. J Educ Psychol 14:193–208

Russell TA (1968) Speed and accuracy of performance in relation to personality in children. PhD thesis, University of London

Shackel B (1967) Electro-occulography. In: Venables P, Martin I (eds) Manual of Psychophysiological Methods. New Holland Publishing Co, Amsterdam

Slater P (1938) Speed of work in intelligence tests. Br J Psychol 29:55–68

Smith EE (1968) Choice reaction time: an analysis of the major theoretical positions. Psychol Bull 69:77–110

Smith J, Jones D, Elithorn A (undated) The Perceptual Maze Test. Interim Report (private circulation) Medical Research Council, UK

Spearman CE (1904a) The proof and measurement of association between two things. Am J Psychol 15:72–101

Spearman CE (1904b) 'General intelligence' objectively determined and measured. Am J Psychol 15:201–293

Spearman CE (1927) The abilities of man. Macmillan, London

Spearman CE, Jones LW (1950) Human ability. Macmillan, London

Spielman I (1963) The relation between personality and the frequency and duration of rest pauses during massed practice. PhD thesis, University of London

Sutherland JD (1934) The speed factor in intelligent reactions. Br J Psychol 24:276–294

Tate MW (1948) Individual differences in speed of response in mental tests of varying degrees of difficulty. Educ Psych Meas 8:353–374

Tate MW (1950) Notes on the measurement of mental speed. J Educ Psychol 41:219–236

Terman LM, Merrill MA (1960) Stanford-Binet intelligence scale: manual for the third revision, form L–M. Houghton Mifflin, Boston

Thorndike EL (1903) Educational Psychology. Lemke and Beuchner, New York

Thorndike EL, Bregman EO, Cobb MV, Woodyard E (1927) The measurement of intelligence. Teachers College of Columbia University, New York

Thurstone LL (1928) The absolute zero in intelligence measurement. Psychol Rev 35:175–197

Thurstone LL (1937) Ability, motivation and speed. Psychometrika 2:249–254

Van der Ven HGS (1971) Time limit tests: a critical evaluation. Ned Tijdschr Psychol 26:580–591

Van der Ven HGS (1974) The correlation between speed and precision in time-limit tests. Ned Tijdschr Psychol 29:447–456

Vernon P (1960) Intelligence and attainment tests. University of London Press, London

Vernon P (1961) The structure of human abilities. Methuen, London

Walsh DM, Walsh MD (1978) Relationship between extraversion and neuroticism, and intelligence for students in grade nine English and mathematics. Psychol Rep 43:15–19

Wechsler D (1958) The measurement and appraisal of adult intelligence. Balliere Tindall Cox, New York

Welford AT (1969) Age and skill; motor, intellectual and social. In: Welford AT, Birren JE (eds) Decision making and age. Interdisciplinary topics in gerontology, vol 4. Karger, Basel

White PO (1973a) A mathematical model for individual differences in problem solving. In: Elithorn A, Jones D (eds) Artificial and human thinking. Elsevier/North-Holland Biomedical Press, Amsterdam

White PO (1973b) Individual differences in speed, accuracy and persistence: a mathematical model for problem solving. In: Eysenck HJ (ed) The measurement of human intelligence. TP Press, Lancaster, pp 246–260

Wolf TH (1973) Alfred Binet. University of Chicago Press, Chicago

Wolff PH (1967) The role of biological rhythms in early psychological development. Bull Menninger Clin 31:197–218

3 Some Major Components in General Intelligence

P.O. White

Introduction

In most cognitive tests of conventional design the subject gains a mark for every problem correctly solved within a time limit. The score thus gained depends in part on the choice of problems attempted, in part on the rate at which the subject works, and in part on the extent to which he abandons problems which, given greater persistence, he might eventually solve correctly. Furthermore, the extent to which these different aspects of test and performance influence total score is quite unknown. Clearly, such a single score can be only an incomplete and probably quite inadequate summary of a very complicated problem solving performance.

This criticism is of course not new. Thorndike et al. (1927) noted that the time spent by a subject on a customary test was a mixture of the time spent in doing some tasks correctly, the time spent in doing other tasks incorrectly, and the time spent in inspecting other tasks and deciding not to attempt them. They argued that while this confusion may be permissible in the practical work of obtaining a rough measure of intellect at small expense of labour and skill, for theory at present and for possible improvement of practice in the future we need to separate the speed of successes from the speed of failures. Thorndike (1925) had argued that ability should be analysed in terms of level, range, and speed and that it seemed desirable to treat these three factors separately and to know the exact amount of weight given to each when we combine them. Eysenck (1973) has pointed out that Thorndike failed to follow up the implications of his penetrating critique.

Thurstone (1937) considered the relationships between ability, maturation, speed, and problem difficulty. He regarded the ability of a subject to perform a specified kind of task as the difficulty at which the probability is one half that he will do the task in infinite time. He suggested that the difficulty of a problem could be assessed in terms of the percentage of errors yielded in a standard group of subjects and he outlined a procedure for estimating a subject's ability, using response times to problems of differing difficulty. Thurstone, too, does not seem to have followed his theoretical contribution with a programme of related empirical work.

There is, of course, a very large literature on the nature and meaning of intelligence, ability, speed, accuracy, and problem difficulty. Most of this literature, both theoretical and empirical, has been covered in two massive, thorough and very critical reviews by Brierley (1969) and by Berger (1976).

In 1948 HJ Eysenck initiated a series of investigations which has led to some clarification of the problems which are involved in making an adequate assessment of cognitive abilities (Furneaux 1952, 1955, 1960; White 1973a, 1973b, 1976; Wiseman 1975, Berger 1976).

Furneaux, whose work has clear roots in the work of both Thorndike and Thurstone, devised a simple conceptual model which illustrated quite clearly the possible ambiguity of the traditional total score and which made it quite clear that in general the basic unit of observation must be the response of the individual subject to the particular problem. It led as well to a number of predictions which in an extensive series of investigations he was able to test empirically. This conceptual model also led White to devise a statistical, latent trait model which incorporates latent ability variables (such as speed and accuracy) and latent continuance variables (such as persistence) for each subject and a vector of problem parameters (such as difficulty level and discriminating power) for each problem. Berger, in a well-conceived and flawlessly executed study, attempted to replicate the main findings of Furneaux's work. He was extremely critical of the results but we feel that his attempt was a resounding success. Wiseman, following the basic logic of White's latent trait model, devised a latent trait model for application to data from traditional time-limit tests. His model included 'ability', 'speed', and 'omissiveness' variables for each subject and 'difficulty level' and 'discriminating power' parameters for each problem.

Since then, White has made considerable advances both in terms of computational efficiency and model simplification and has now fitted his model to data from four tests of cognitive ability on a large sample of school children of both sexes.

In our opening paragraph we paraphrased the main argument against the global score so characteristic of the traditional tests of cognitive abilities. Today the argument seems self-evident.

Yet today, more than 30 years after Furneaux's first talk on the subject, its impact on published tests has been confined to the Nufferno tests, which he himself designed and published.

The form of our outline of the Eysenck-Furneaux notions has been influenced by three important facts. In the first place there is very little published material on the matter. We have seen but three reports by Furneaux and one of these is essentially a promotional blurb on the Nufferno tests. We have seen but three reports by Eysenck on the matters involved (1953, 1967, 1973). These are clear, characteristically terse, simple expositions of the problem. They are a delight to read. They do not, however, and of course were not meant to, serve as definitive documentation for the model.

A second fact which has influenced our mode of presentation is that now, more than a quarter of a century later, it is not at all clear which parts are due to whom.

Finally, we have had considerable difficulty in coping with both the 1952 paper and the 1960 paper by Furneaux. The former is but one page in length and has two erroneous formulae, which were corrected in an erratum published in a later issue of the same journal. Brierley (1967) sensed that there was something wrong with the third formula and suggested some modifications, which seem only to have added to the confusion. In fact, though the intent of the formula is quite clear we were unable to resolve the matter satisfactorily. In the same paper Furneaux states that the equation has the same form as expressions suggested previously by Thorndike and by Thurstone. We have not yet seen support for this claim. Quite recently, Furneaux (personal communication), when queried on the matter, stated that there was not much point in worrying about such specific details since these formulae had been superseded anyway. Since then we tracked down some unpublished notes dated 1953. There, too, the disputed formula has essentially the same form.

His 1960 chapter is generally regarded to be his definitive work on the subject. We found it, too, rather puzzling. In his introductory comments he states that its function is to sketch in the background of results and ideas. Some 12 pages later he apologizes for the amount of detail given. [Brierley (1969) devoted a five-page appendix to a listing of Furneaux's symbols!] We found his notation extremely cumbersome and his casual introduction of new symbols and of abbreviations exasperating. We found his formal statement of the conceptual model to be incomplete in the sense that the notion of the comparator device was introduced, almost as an afterthought, after his discussion of his results, and we found his outline of the statistical analysis extremely difficult to follow. Nevertheless, we regard this paper as one of the more important original documents yet published on the measurement of cognitive abilities. We doubt seriously that anyone would benefit from a thorough, nit-picking documentation of all its blemishes. Instead, we emphasize its more important aspects.

We know Furneaux as a thorough, dedicated, and conscientious worker. We found him unstinting both with respect to time and patience in trying to answer our virtually continuous barrage of questions regarding his conceptual model and the details of his statistical analyses. We present instead our own statement of his conceptual model as we came to know it over a quite extended period, partly through struggling with the published material, but mainly through his own patient attempts to explain what he meant and to try to sort out misunderstandings regarding the published work.

The Eysenck-Furneaux Paradigm: Part I

Let us now come back to our basic problem of what happens when a subject is confronted with a traditional time-limit test of cognitive ability. He is given the instructions and is guided through one or more sample problems. He is given the signal to 'go' and after a period of time is told to 'stop'.

Typically, much goes on during this period and not all of it is relevant to what the tester is trying to measure. This includes, for example, such things as turning pages, writing answers, crossing out wrong answers, backchecking, replacing broken pencils, and perhaps looking around to see how others are doing. Quite clearly, even if we ignore important aspects like correct versus incorrect and abandon versus non-abandon, the 'time taken' does not tell us even for two subjects who complete the test in the allotted time how much time was taken in solving the problems per se and how much time was spent in extraneous activity.

Even if we know the total time taken on relevant problem solving activity we are not too much better off because two subjects with the same time spent on problem solving may well have worked on different numbers of problems.

Furthermore, two subjects who worked on the same number of problems and spent the same amount of time working on them may well have distributed their time differently. More likely, of course, is the possibility that they would have worked on different numbers of problems, they would have spent different amounts of time working on the problems, and they would have distributed their time differently as well.

If we now consider a group of subjects all faced with the same set of problems and remove the time limit so that each subject reaches the last problem and if we assume that we know for each subject how much time he has spent on problem solving, we still have considerable ambiguity. And this is true even if we assume that all subjects in our group have achieved the same number of correct answers.

For example, a typical pattern is for the subject to abandon the more difficult problems, to pass the easier problems, and to fail the intermediate ones. But many subjects do not perform according to this neat scheme. Some fail problems which are easier than some of those which they have passed. Others abandon problems which are easier than others which they have passed. Still others pass problems which on the basis of their overall performance we would expect them to fail. Probably this is because they guess rather than abandon and happen to be lucky. Some subjects spend very little time on some problems before abandonment and on other problems they spend a considerable time before abandonment. Perhaps they decide almost instantly to skip the problem and to try to pick up the lost points on the more obviously easy problems.

Furneaux placed great emphasis on the

distinction we have just made regading different types of abandonment. He also placed great emphasis on the possibility that a subject both slow and lacking in persistence might tend to give up the more difficult problems before he has given himself time to reach a solution or alternatively to record an answer hastily guessed and inadequately checked. In recognition of the fact that 'persistence' is not the only determinant of abandonment he suggested that it might be more reasonable to use the word 'continuance', which he felt had no 'aetiological presuppositions'.

Furneaux felt that in order to lay the foundations of the study of problem solving behaviour we should concentrate at first on a few limited fields of study and that thus, in the initial stages at least, we must ignore many attributes which undoubtedly influence problem solving responses.

At the observational level this decision means that we record for each subject, on every problem, whether or not he abandons the problem and whether or not he solves the problem correctly. We record as well, for each subject, on every problem, his response time or latency. We thus have three response categories – right (R), wrong (W), and abandon (A). We emphasize that we are grouping together into category A not only those problems which the subject abandons because his continuance is exhausted but possibly, for some subjects at least, problems which he has abandoned for other reasons such as, for example, strategic abandonment. We emphasize as well that categories R and W may well, for same subjects at least, include problems on which the subject has 'guessed' and on which the outcome is more or less random dependent, of course, on whether, and if so to what extent, the subject has made use of partial knowledge.

We emphasize also that we ignore the many possible determinants of outcome and characterize each subject in terms of three attributes only. These are: (a) the rate at which a 'search' process, having as its object the evolution of a solution, proceeds – we call this attribute 'mental speed'; (b) the efficiency of those cerebral mechanisms whose function it is to check the adequacy of trial solutions, as they arise, against the demands of the actual problem – we call this attribute 'accuracy' (or 'solution recognition efficiency'); (c) the 'continuance' displayed by a person in the face of the discouragement resulting from failure to find a solution which can be accepted as adequate. Finally we emphasize that we ignore specific aspects of the different problems and differentiate among problems only in terms of their respective 'difficulty levels'.

We now turn to an outline of Furneaux's conceptual model – his analysis of a formalized problem solver having only such characteristics as are explicitly assigned to it.

Furneaux proposed that, initially, we regard our formalized problem solver as a 'black box' device or problem box containing an unspecified mechanism of such a nature that when it is supplied with an input in the form of a problem an output results. Associated with each output is a 'completion time', t. This is the elapsed time between the feeding in of the input and the production of the output.

When an output is produced it will correspond to one of three mutually exclusive outcome categories. These are R (correct response or right answer), W (error response or wrong answer), or A (abandonment). We inspect each output and label it R, W, or A depending on its outcome category and we label its associated completion time, t_c, accordingly. Thus, each completion time becomes one of t_R (time to correct response), t_W (time to error), or t_A (time to abandonment). And these, too, are mutually exclusive.

Thus far, our hypothetical problem box is rather empty. It merely mirrors the observational scheme which we outlined in the previous section. It receives inputs and it produces outputs and completion times which we observe, classify, and label. We have not given it any capacity to distinguish among inputs and we have specified nothing whatsoever about the mechanism which determines, for a specified input, either the resultant outcome or the production time associated with the outcome.

Suppose, now, that we feed into our formalized problem solver a very large number of equivalent inputs, that we observe all outputs and completion times, that we allocate each output to the appropriate outcome category (R, W, or A), and that we label its associated completion time, accordingly, as one of t_R, t_W, or t_A. We are now in a position to compute, for this particular formalized problem solver on this particular group of equivalent, but otherwise unspecified, problems, a variety of summary statistics. We list but a few of the many possibilities: (a) proportion right $\left(\frac{R}{R+W+A}\right)$, (b) proportion wrong $\left(\frac{W}{R+W+A}\right)$, (c) proportion abandoned $\left(\frac{A}{R+W+A}\right)$, (d) mean time to correct response ($\bar{t}_R$), (e) mean time to abandonment ($\bar{t}_A$), and (f) mean time to non-abandonment ($\bar{t}_{RW}$). These are but a few selected examples. Other possibilities include variances, covariances, correlations, measures of range, regression coefficients, estimates of skewness, or estimates of kurtosis.

But what, if anything, have we gained by feeding this large set of equivalent, but otherwise unspecified, group of problems into our formalized, but otherwise unspecified, problem solving device, by observing and categorizing the outputs and their corresponding production times, and by summarizing the resultant data in a set of arbitrary statistics? All we have done, of course, is to add to our specified observational scheme an additional data reduction scheme. And this, of course, is not of much help to us since, with a still unspecified mechanism within our formalized problem solving device, we have no guidance as to how we might interpret these statistics in order to infer any properties of the group of problems which were processed or any properties of the problem solving device itself.

We now introduce some formal structure into our problem solving device. Let us assume, first, that upon input of a problem a timer is initialized and that it subsequently begins to count successive time intervals, that it continues to do so until an output is produced, and that when this occurs its value is preserved. This value, of course, is the completion time, t_c, to which we have already referred. Let us assume further that our formalized problem solver has been fitted with a 'time switch', which operates according to the following rules:

1. If the elapsed time since input of the problem is less than some lower limit t_l, it will not operate.
2. If the elapsed time since input of the problem reaches some upper limit t_u, it will operate with certainty and will enforce an output with associated completion time $t_c = t_u$. In this case the associated outcome is an abandonment and the completion time associated with the outcome becomes an abandonment time, $t_A = t_u$.
3. If the elapsed time since input of the problem has reached t_l but has not yet reached t_u, the time switch will operate on a purely random basis with the probability of operation constant within the interval $t_l - t_u$.
4. All abandonment outcomes are associated with outputs forced by the time switch. Given these four rules, it follows directly that:
   1. All outcomes associated with completion times less than t_l are either correct responses or errors.
   2. All correct responses or errors are associated with completion times less than t_u.
   3. All abandonment times are equal to or greater than t_l.
   4. No abandonment time is greater than t_u.
   5. Any outcome (correct response, error, or abandonment) may be associated with a completion time t_c in the interval $t_l \leqq t_c \leqq t_u$.

Our formalized problem solver now has, in addition to its input output and timing capabilities, an attribute corresponding to the concept of 'continuance'. We have not yet endowed it with any mechanisms corresponding to the concepts of 'mental speed' or 'accuracy'.

Yet even in this rudimentary form with no details at all specified regarding mechanisms concerned with mental speed or accuracy, it is quite clear that the continuance characteristics of our device (t_l and t_u) can have a profound effect, in certain circumstances at least, on the relative frequencies of the outcomes, on the distributions of their associated completion times, and on statistics derived from these values.

We will discuss this last point in considerable detail later. Before doing this, however, we introduce the mechanisms which give our hypothetical problem solving device attributes corresponding to the concepts of 'mental speed' and 'accuracy'.

When a problem is fed into the device, activity is initiated and this activity continues until an output is produced. We use the term 'search' to denote this activity. For the formal analysis it is not necessary to go into specific details of the searching activity. We postulate only that each step consists of the retrieval of a potential solution to the problem (or of some set of elements corresponding to a potential solution to the problem and that this potential solution to the problem is examined to determine whether it constitutes the solution to the problem. We call the mechanism which compares the potential solution with the problem and determines whether the potential solution consitutes the solution to the problem 'the comparator'. We assume that if the potential solution does constitute the solution to the problem it will be accepted as such with probability 1 but that potential solutions which do not consitute the solution to the problem will be accepted as being correct with probability $1-\varepsilon$. Following Furneaux we refer to ε as the 'solution recognition efficiency' or 'accuracy' although it clearly corresponds to non-solution rejection. The rate at which the 'search' activity proceeds we call the 'speed' of the device. Hypothetical problem solvers may thus differ in any of three respects: (a) they may differ in the speed at which the 'search' activity proceeds, (b) they may differ in the accuracy with which incorrect solution are rejected, and (c) they may differ in their values of t_l and t_u. In other words, hypothetical problem solvers may differ only with respect to 'speed', 'accuracy', and 'continuance'.

This completes our formal definition of the conceptual model which Furneaux has proposed. It is, as he well knew, and so strongly emphasized, a greatly over-simplified device. And yet, as he showed, even this simplistic device turns out to be extremely complex when we come to examine its input-output relationships.

The Eysenck-Furneaux Paradigm: Part II

In order to illustrate the complexity implied by this simplistic formulation we examine in some detail a number of specially chosen cases. Each case has been chosen for its apparent simplicity. In each case we utilize an argument of the form 'other things being equal'.

Let us consider, first, a situation in which we have two hypothetical problem solving devices which differ only in accuracy. They are identical both in terms of their 'speed' (i.e. in the rate at which the search activity proceeds) and of their 'continuance' (i.e. in the particular values of t_l and t_u at which their respective time switches are set). They differ only in that one (the more accurate device) is more likely to reject a potential solution which is incorrect than is the other (the less accurate) device.

Let us assume that a set of problems which span a considerable range of difficulty has been fed into each device, and that the outputs and their associated completion times have been recorded. Clearly, since our two devices differ neither in terms of speed nor in continuance we expect the proportion of abandonments to be the same and we expect the distribution of abandonment times to be the same as well. It is clear as well that we should expect the more accurate device to obtain more correct responses and thus fewer errors than the less accurate device. It is also clear that we should expect the mean difficulty of problems correctly solved to be higher for the more accurate device than for the less accurate device. Since we have assumed that the search activity proceeds at the same rate for both devices this implies that the more accurate device will have spent more time working on problems for which it eventually obtained the correct response than will the less accurate device. This is a direct consequence of the fact that in this formulation 'mental speed' is essentially the regression of problem difficulty on solution time. It follows that under our hypothesis of 'equal speed equal continuance' we should expect the mean time to correct response to be higher for the more accurate device than for the less accurate device. The 'obvious' conclusion that the former device is slower than the latter device is not valid since by hypothesis the two devices do not differ in speed.

Let us now consider a second example. We have the same set of problems but this time we have two problem solving devices which are identical with respect to both 'mental speed' and 'accuracy'. They differ only in that the second device has less 'continuance' than does the first device. Thus for the second device the time switch is set such that the values of t_l and t_u are considerably lower than are the corresponding values for the first device. It is clear from our description of the hypothetical mechanism that we should expect the second device to produce more outputs in the category 'abandon' and that we should expect mean time to abandonment to be higher for the first device. Thus we expect number right plus number wrong to be higher for the first device than for the second device since number right plus number wrong plus number abandoned is the same for both devices. We should expect that number right/(number right + number wrong) would be the same for both devices.

For both devices we should expect that a proportion of the outputs of type 'correct response' and 'error response' will be lost due to the intervention of the time switch. However, we should expect this proportion to be higher for the second device due to the intervention of the time switch at shorter times. It follows that we should expect 'mean time to correct response' and 'mean time to error' to be lower for the second device than for the first. The 'obvious' conclusion that the second device is faster than the first because it achieves both correct solutions and errors in less time is not valid since by hypothesis the two devices do not differ with respect to speed.

Our third 'simple' example parallels the first two. Again, the same set of problems is input into two hypothetical problem solving devices which differ in but a single way. This time both devices have identical 'accuracy' and 'continuance' characteristics. They differ only in that the second device is slower than the first. That is, the speed of the 'search' activity is greater for the first device than for the second. Given that the two devices have operated for some time t without the production of an output it is more probable for the first device that the 'potential solution' about to be examined by the comparator constitutes the 'correct solution' to the problem. The complication here is that in this time period the first device will have made more comparisons than the second and since both devices have by hypothesis the same probability of accepting an incorrect solution as being correct on any given comparison the first device will have had more opportunity of producing an error by this time. Clearly, the probability of a correct response by time t will be a function of the difficulty of the problem and of the speed-accuracy characteristics of the particular device. It seems clear that if the first device produces, for a particular problem, an output of type 'correct response' at time t_1 and the second device (the slower one), for the same problem, produces an output of type 'correct response' at time t_2, we should expect t_1 to be less than t_2.

Furneaux argued that for each problem solving device there exist two critical values of problem difficulty. For problems with difficulty levels lower than the lower critical value d_l, solution times will be less than t_l, the time switch will never intervene, and all outputs associated with such inputs will be of type 'correct response' or 'error'. He argued that statistics based on these outputs would be unambiguous in the sense that they would reflect the 'speed-accuracy' characteristics of the device but not its 'continuance' characteristics. He called such inputs 'unambiguous inputs'. He then argued that for all problems with difficulty level above some upper critical value d_u the required solution time (i.e. for 'correct response' or 'error') would be greater than t_u. However, for such problems the time switch would always intervene before the emergence of such outputs and thus for these problems all outputs would be of type 'abandonment'. For problems with difficulty level in this region the statistics associated with the outputs would also be unambiguous in the sense that they would reflect only the 'continuance' characteristics of the device and not its 'speed accuracy' characteristics. He called inputs with difficulty level in this region unambiguous as well. He then argued that for inputs with difficulty level in the interval $d_l - d_u$ the number of outputs of type 'correct response' or 'error' will be reduced due to the intervention of the time switch. Thus for problems with difficulty levels in this region the statistics associated with the respective outcomes will be affected both by the 'continuance' characteristics of the device and by its 'speed-accuracy' characteristics. We have seen in our three examples above the sort of interactions which may occur.

Furneaux emphasized strongly that the values of d_l and d_u will differ from one problem solving device to another and that their values will depend intimately on its 'speed-accuracy-continuance' characteristics.

Furneaux argued that these conclusions, based on the analysis of a formalized problem solving device, must be applied equally to the human subject. He gave an example which illustrates clearly some of the implications of his analysis for test construction. He considered a test made up of problems having such low difficulty values that no individual in the population returns any incorrect solutions. As the items are 'easy' they will all be solved quite quickly, so continuance will not be a determinant of success. Thus, under these circumstances the only factor influencing score if the test is untimed is problem solving speed. He assumed that following a set of n_e such easy problems subjects proceed immediately to a set of n_m problems at a higher level of difficulty. He assumed that problems in this set are sufficiently 'difficult' to lead to the production of some incorrect solutions by most members of the group but that none are so difficult in relation to the range of mental speed within the group as to be given up as insoluble. He pointed out that if this two-part test is administered with time limits, the slowest members of the group will still be working on the easy problems when the time limit expires and their scores will be determined by problem solving speed. The moderately fast subjects will all complete the easy problems but at varying rates and when the test finishes they will have been working for different times on the problems of moderate difficulty. He points out that during the time an individual is working on the moderately difficult problems the rate at which his score will increase will be in part a function of his problem solving speed but also in part on his tendency to produce incorrect solutions (i.e. on his accuracy). Thus, he concluded that for this group of moderately fast individuals the final score attained will be determined in a fairly complex fashion both by speed and by accuracy. The very fast members will all finish the test but will be distributed in terms of their tendencies of produce errors (i.e. in terms of differences in accuracy). He then assumed that we add a further set of n_d very difficult problems to the test and that the subjects proceed to work on these as well. He noted that the final score for the fast members of the group, who reach these problems within the time allotted, will depend on 'continuance', for those who are less continuant will lose possible increments of score through abandoning their efforts to obtain a solution before sufficient time has elapsed for a solution to emerge. He argued that if such a test is administered with time limits to a fairly homogeneous group it could measure mainly 'speed', mainly 'accuracy', or mainly 'continuance' depending on the interaction of the range of ability represented in the group, the time allowed for the test, and the numbers of problems in the three groups. He noted that in a heterogeneous group the attributes measured could, under some conditions, vary all the way from pure 'speed', for some individuals, through various combinations of 'speed', 'accuracy', and 'continuance' for other individuals. He argued that if such a test is administered without time limit the manner in which these factors combine will be modified but that the same types of complication will arise as when a time limit is imposed. He suggested that 'such a test cannot be said to measure any single, clearly defined trait, and that under some circumstances the same tests will be comparing different subjects in terms of quite unrelated attributes.'

In his 1952 paper Furneaux reported, very briefly, on the outcome of the first 3 years of the Eysenck-Furneaux collaboration. All of this work involved problems of the type included in the well-known letter series tests devised by Thurstone. He gave three equations. The first two involved the normal probability integral and in both he inadvertantly omitted to include a rather vital t. He corrected these errors in an erratum which appeared in a later issue of the same journal. The first of these equations expressed the probability of 'correct response by time t 'as a function of the difficulty of the problem and of the speed of the subject. The second of these equations expressed the probability of 'abandonment by time t' as a function of the continuance of the subject. Both equations, of course, involved the variable t. Direct application of a well-known result in the calculus to the corrected equations leads to the following equivalent statements. (a) The logarithm of 'time to correct response' is distributed normally with expectation $(mD-S)$ and variance σ_S^2 and (b) the logarithm of 'time to abandonment' is distributed normally with expectation C and variance σ_C^2. The problem difficulty, D, is defined in the conventional, but arbitrary, manner as the percentage of an unselected adult British population who fail the problem, and S and C are constants which define the subject's 'speed' and 'continuance'. He stated that '...there is very strong evidence that m and σ_S are identical for all subjects who have so far been examined, not only within a particular test situation, but also between situations involving different degrees of motivation'. He reports an estimate of $m = .013$ with an estimated standard error of .0005 and he reports an estimate of $\sigma_S = .12$ with an estimated standard error of .005. He suggests that 'It seems probable that these two parameters can join Philpott's fluctuation constants... bringing the number of true psychological constants up to four'. In the 1960 chapter he reports a similar finding. In this analysis problem difficulty is defined in terms of mean time to correct response; yet, once more, his estimates of m and of the corresponding variance are constant across subjects and, in all cases, m is very close to unity. Here, though, he makes it clear that he tried to choose the scale so as to give m the convenient value of 1.0 and that he introduced the transformation from raw time to log time as a variance-stabilizing device. He seems to have been very successful on both counts. We note that if, as in his first analysis, D is defined as percentage of errors and $m = .01$, then mD is the proportion of errors. If we define

D as a proportion in the first place then m becomes unity as in the later analysis, which uses a time-based difficulty scale. The implication seems obvious.

Furneaux notes towards the end of the 1960 chapter that he had also attempted to construct difficulty scales based on mean time to error. Here, though, because mean time to error is less than mean time to correct response at any level of difficulty, because this difference increases with increasing difficulty, because the variance of error times is higher than that for time to correct response at any level of difficulty, and because this difference, too, increases with increasing difficulty, his efforts were not very successful. He notes that these discrepancies are precisely those which would be predicted from his conceptual model. He notes further that, at the low values of difficulty required for unambiguous outputs, there are fewer errors produced and thus the mean time to error and variance of mean time to error are both measured less accurately. This too, he argues, is in line with his conceptual model.

He notes as well that his attempt to construct a difficulty scale based on proportion of errors was somewhat more successful. This scale correlated .92 with that based on time to correct response.

We now return to the third equation in the 1952 paper. This equation (which we do not reproduce) expresses the probability of an error response as an increasing, monotonic function of the difficulty level of the problem and as a decreasing, monotonic function of the subject's 'solution recognition efficiency' or 'accuracy'. Indeed, the probability of an error is an increasing monotonic function of $(D-E)$, where D, as before, is problem difficulty and E is the subject's solution recognition efficiency, or accuracy. He states that both Throndike and Thurstone proposed functions of the same form but we have not yet been able to find their equations in the literature. We wrote a program to evaluate the integral, which hè gave, using numerical quadrature and we found that it is quite easy to generate error probabilities in excess of unity. On the advice of Furneaux (personal communication) we abandoned our persistent attempts to sort out the matter.

In neither the 1952 paper nor the 1960 chapter does Furneaux give details regarding the 'accuracy' scale. It seems quite clear, however, that it must be based on (number right)/(number right + number wrong).

He reports correlations among his final speed, accuracy, and continuance scales. They are as follows: .38 (speed-accuracy), .27 (speed-continuance), and .31 (accuracy-continuance). He concludes, and we agree, that the three scales are relatively independent and that they demand separate consideration.

We note that this set of correlations is quite consistent with a factor analytic model with a single general factor on which both speed and accuracy have loadings of about .62 and continuance has a loading of about .45. This result implies that, if we take the general factor as criterion and regard the speed, accuracy, and continuance scales as independent variables in a linear regression equation, the multiple correlation is about .77. Letter series problems of the type used in the Eysenck-Furneaux investigations are generally presumed to define the primary mental ability of induction, I. Thus, it could be argued that in the simple regression procedure just described, we have in some sense broken down this induction factor into three relatively independent components derived from a careful, psychological analysis of the problem solving process. If we accept this proposition, it follows that, in principle, each of the other primary mental abilities could be similarly broken down into their constituent parts. We would then be in a position where we could test relatively precise hypotheses regarding the generality of the speed, accuracy, and continuance components and of their possible relationships with more general factors like 'general intelligence', 'fluid ability', and 'crystallized ability'. This is, indeed, the direction our own empirical work with Eysenck has taken. Here, though, we are not concerned in detail with these substantive matters nor with their implications for a theory of intelligence. Instead, we turn to our own attempts to construct a latent trait model which incorporates latent ability variables (such as mental speed and accuracy) and latent continuance variables (such as continuance or persistence) for each subject as well as a vector of parameters (such as difficulty level) for each problem.

Table 1. The regression hypothesis

(1) $ln\, T_s \sim N[(mD-S), \sigma_s^2]$

(2) $m=.01$

(3) $D=100-R$

(4) $R=$percentage correct

(5) $mD=$proportion incorrect$=ln\, d$

(6) $S=ln\, s$

(7) $E[T_s]=\exp[(mD-s)+\frac{1}{2}\sigma_s^2]$

$$=\frac{d}{s}\exp[\tfrac{1}{2}\sigma_s^2]$$

(8) $E[s\, T_s]=d$

where $s \leftarrow \dfrac{s}{\exp[\frac{1}{2}\sigma_s^2]}$

A Probabilistic Latent Trait Model

Our own attack on the problem began with three observations:

1. In the Furneaux model, 'speed' is the regression of problem difficulty on time to correct response. In Table 1 we sketch the development. Equations (1)–(4), as we have already seen, follow directly from the first equation in the 1952 paper. In Eqs. (5) and (6) we introduce a change of variable for both problem difficulty (*d*) and speed (*s*). Thus, our problem difficulty, *d*, is an exponential function of Furneaux's difficulty, *mD* (proportion of failures) and our speed, *S*, is an exponential function of Furneaux's speed, *S*. Equation (7) follows from a standard, and well-known, result in distribution theory (Parzen 1960). We note that σ_S^2 is, for a particular subject, constant across problems and that it is also constant across subjects. In Eq. (8) we absorb this constant into *s*. We use the replacement operator $\leftarrow$ rather than the equivalence operator $=$. Thus, expected time to correct response, on a particular problem, is shorter for a fast subject than for a slow subject and, for a particular subject, expected time to correct response is proportional to problem difficulty.
2. The expected time to abandonment on a particular problem is a decreasing monotonic function of the subject's continuance. It follows directly that the probability of an abandonment by time *t* is also a decreasing montonic function of the subject's continuance. This observation follows from the second equation in the 1952 paper, which we have already discussed.
3. The probability of a correct response, on a particular problem, given completion and non-abandonment, is an increasing monotonic function of the subject's solution-recognition efficiency or accuracy and is, for a particular subject, a decreasing monotonic function of the problem difficulty. This observation follows from the third equation in the 1952 paper. The condition on non-abandonment follows clearly from the fact that Furneaux attempted to base the scaling of difficulty and of speed and accuracy only on statistics derived from 'unambiguous inputs'.

On the basis of these observations, which come directly from Furneaux's analysis, we wrote the five equations displayed in Table 2. They included two latent ability variables (speed, *s*, and accuracy, *a*) and a latent continuance variable, *p*, for the subject and they included a problem parameter (difficulty level, *d*) for the problem. They also included completion time, *T*.

It seemed clear that these five equations incorporated all of the relationships outlined in the three observations above and that they incorporated much, if not most,

Table 2. A very simple latent trait model

(1)	P[correct | not abandon, completion at time T] $=\alpha$
(2)	P[abandon | completion at time T] $=\beta$
(3)	$\ln\frac{\alpha}{1-\alpha}=a-d$
(4)	$ln\frac{\beta}{1-\beta}=T-p$
(5)	$E[sT$|correct, completion at time $T]$ $=d$

Table 3. A less simple latent trait model

(1)	P[correct | not abandon, completion at time T] $=\alpha$
(2)	P[abandon | completion at time T] $=\beta$
(3)	$ln\frac{\alpha}{1-\alpha}=a+sT-d$
(4)	$ln\frac{\beta}{1-\beta}=T-p$

of the essence of the conceptual model proposed by Furneaux. However, we felt that, 'other things being equal' (that is, given equal accuracy and equal persistence) if two subjects worked on the same problem for the same time the faster subject, the one with the faster search activity, should have a higher probability of a correct response and we tried to incorporate this notion into Eq. (3). It then became clear that the assumption implied by Eq. (5) was no lonser necessary. This new version of the model included the equations displayed in Table 3. No new terms were introduced, one equation was dropped, and the new version of Eq. (2) seemed to reflect our intent very well.

However, it was now clear that the curve relating β to T would have the same shape for all subjects and would just shift up or down as p varied from subject to subject. In the Furneaux model, the probability of an abandonment is zero for all completion times less than t_l and it is unity for all completion times equal to or greater than t_u.

Table 4. A more elaborate latent trait model

(1)	P[correct | not abandon, completion at time T] $=\alpha$
(2)	P[abandon | completion at time T] $=\beta$
(3)	$ln\frac{\alpha}{1-\alpha}=D(a+sT-d)$
(4)	$ln\frac{\beta}{1-\beta}=c(T-p)$
(5)	$c, D>0$

In the interval t_l-t_u it increases linearly and the rate at which it increases is inversely proportional to the length of the interval. Our familiarity with the latent trait models of Rasch and with the two-parameter Birnbaum model made it clear that we could incorporate the same notion into Eq. (4) by adding a latent continuance variable, c, for each subject. At the same time we decided to add a second problem parameter D, analagous to discriminating power in the two-parameter Birnbaum model, and to call it, by analogy, discrimination power. The new equations, displayed in Table 4 now numbered five. The constraints imposed by Eq. (5) are purely technical. They make the equations function as we intended and not the opposite way around. We note that if we set $C=D=1$ the equations reduce to those in Table 3. Thus, the model represented in Table 3 is a special case of that represented in Table 4.

In this form our model incorporates two latent ability variables (mental speed and accuracy) and two latent continuance variables (persistence, and a shape constant for the continuance function) for each subject and two problem parameters (difficulty level and discriminating power) for each problem.

The probability of a correct response given non-abandonment and completion time T is a cumulative logistic function[1] of $D(a+sT-d)$. If we write this as $\alpha=\Psi[D(\theta-d)]$ with $\theta=a+sT$, we see its relation with the

1 $\Psi[\cdot]=\{1+\exp[-(\cdot)]\}^{-1}$

two-parameter Birnbaum model in which θ is the unidimensional latent ability variable. We replace this with a time-dependent function involving the two latent ability variables 'speed' and 'accuracy'.

The probability of an abandonment is, similarly, a cumulative losistic function of $c(T-p)$. The form is clearly the same as that of the two-parameter Birnbaum model but here we have a time-dependent function of the two latent continuance variables for the subject.

It we set $c=D=1$ then α is a cumulative logistic function of $(\theta-d)$ and β is a cumulative logistic function of $(T-p)$. The similarity in form with respect to the simple logistic model of Rasch (1960a, b) is obvious.

We now build up the model on a more formal basis and in a much more general form. The models outlined in Tables 2–4 all turn out to be special cases of the more general model and they all share the 'separability' properties which characterize the general model and which have important implications when we attempt to fit special cases of the general model to sets of empirical data.

We now proceed to derive our model. First we set up a basic framework using as few symbols as possible and ignoring subscripts as much as possible. We restrict ourselves for the moment to an individual subject and to a single problem. We assume that the subject is allowed to take as much time as he needs, that he has been instructed to try to solve the problem as quickly as possible, that he has been cautioned very strongly not to guess, and that he has been provided with the option of proceeding to the next problem without committing himself. And, of course, we assume that he follows our instructions.

We now assume that the problem is presented to the subject and that at the instant of presentation a digital timer begins to count successive time intervals. We assume that the problem solving process continues until, eventually, completion occurs and, simultaneously, the timer is stopped at time T. We say that completion occurs at time T and we note that given completion at time T, one of three events occurs. These are as follows: (a) The subject abandons the problem and proceeds to the next problem without committing himself. He thus forfeits his chance of scoring a correct response but guarantees that he does not score an error. In this case he scores an abandonment. (b) The subject responds by indicating that he is sure that a particular answer is the solution to the problem and he does in fact select the correct answer. In this case he scores a correct response. (c) The subject responds by indicating that he is sure that a particular answer is the solution to the problem but the answer he selects is not the correct one. In this case he scores an error.

We let γ_t, $t=1, 2, \ldots, T$ be the probability of completion at time t, i.e. during the t'th time interval. We let β be the probability of an abandonment given completion at time T and we let α be the probability of a correct response given non-abandonment and completion at time T. We now formalize the above statements in Eq. (1)–(5) of Table 5.

We have just discussed Eqs. (1)–(3). In Eq. (4) we state the obvious fact that no abandonment can occur until completion takes place. In Eq. (5) we state the obvious fact that if an abandonment occurs then a correct response cannot occur.

Table 5. Eliminating γ_T (conditioning on completion time)

(1)	P[correct | not abandon, completion at time T] $=\alpha$
(2)	P[abandon | completion at time T] $=\beta$
(3)	P[completion at time T] $=\gamma_T$
(4)	P[abandon | continuance at time T] $=0$
(5)	P[correct | abandon, completion at time T] $=0$
(6)	P[correct, not abandon, completion at time T] $=\alpha(1-\beta)\gamma_T$
(7)	P[correct, not abandon | completion at time T] $=\alpha(1-\beta)$

We note that in stating our assumptions and in setting up Eqs. (1)–(5) we stated nothing that we had not already stated in words.

And yet, having defined our terms and written our equations, we may manipulate the equations according to the axioms and theorems of probability theory and derive results which are not readily apparent given the purely verbal statement.

We look first at the joint distribution of the outcome 'correct response' and 'time to completion'. Repeated use of the law of compound probabilities[2] leads directly to Eq. (6). Clearly, in order to write a model for this distribution we must specify a functional form for γ_T as well as for α and β.

However, the law of compound probabilities leads directly to Eq. (7), which states the probability of the joint event [correct response, not abandon] conditional on 'completion at time T'. This is the joint distribution of the events 'correct response' and 'non-abandon' conditional on 'completion at time T'. The striking thing about Eq. (7) is that it does not involve γ_T.

It also follows from Eqs. (1)–(5) that P [incorrect, abandon | completion at time T] is equal to β, that P [incorrect, not abandon | completion at time T] is equal to $(1-\alpha)(1-\beta)$. Thus, in all cases the joint distribution of outcomes when we condition the model on completion at time T does not involve γ_T.

This, then, is the conceptual basis of our model. In the following sections we build up the model from scratch, allowing for individual differences among subjects and for differences among problems and we provide basic functional forms (i.e. cumulative logistic) for both α and β. We allow for each problem to be characterized by a vector of latent ability variables and by a vector of latent continuance variables.

We build up a likelihood function for each subject and then we build up a likelihood function over all subjects. We show how the likelihood function factors into independent parts and we describe two basic iterative algorithms for fitting the model to sets of empirical data. We do all this for the general model without specifying problem parameters or latent trait variables explicitly. We thus provide the basic tools needed for working with any one of a whole family of possible models within the framework outlined above.

We assume that a set of n problems has been administered to each of N subjects and that each subject is allowed to take as much time as he needs on any problem, that he tries to solve the problem as quickly as possible, and that he has the option of proceeding to the next problem without committing himself if he feels that he does not know the correct answer. We assume that the subject does not guess and we assume that the outcome in terms of correct response, error response, and abandonment as well as the associated completion time has been recorded for each problem.

We use the subscripts j and k as alternate subscripts to index problems and the subscript i to index subjects. We define two discrete binary random variables X_{ji} and Y_{ji} and the continuous random variable T_{ji}. The random variable X_{ji} takes the value 1 if subject i gives the correct response on problem j and it takes the value 0 otherwise. The random variable Y_{ji} takes the value 1 if subject i abandons problem j and it takes the value 0 otherwise. We use x_{ji} and y_{ji} to denote realizations, respectively, of X_{ji} and Y_{ji}. The random variable T_{ji} is the completion time for subject i on problem j and its realization is t_{ji}.

If subject i gives the correct response to problem j then $x_{ji}=1$, $Y_{ji}=0$, and the completion time, t_{ji}, is the time to correct response. If subject i gives an error response to problem j then $x_{ji}=0$, $y_{ji}=0$, and the completion time, t_{ji}, is the time to error. Finally, if subject i abandons problem j then $x_{ji}=0$, $y_{ji}=1$, and the completion time, t_{ji}, is the time to abandonment. This exhausts the possibilities. If a subject abandons a problem then he forteits his chance of getting a correct response on the problem and

2 The law of compound probabilities:
$P[E_1, E_2]=P[E_1 \mid E_2]\, P[E_2]=P[E_2 \mid E_1]\, P[E_1]$

the response $x_{ji}=1, y_{ji}=1$ is impossible and thus occurs with probability 0.

We assume that each subject is characterized by p latent trait variables and that these consist of p_1 latent ability variables and p_2 latent continuance variables (thus, $p=p_1+p_2$). We assume also that each problem is characterized by a vector of q problem parameters.

We let w_i denote a column vector of order p_1 containing the unobservable latent variables for subject i and we let ϕ_i denote a column vector of order p_2 containing the unobservable latent continuance variables for subject i. Finally, we let v_j denote a column vector of order q containing the unobservable problem parameters for problem j.

We have now defined eight vectors for each subject (X_i, Y_i, T_i, x_i, y_i, t_i, w_i, and ϕ_i) and one vector, v_j, for each problem. We now define matrices X, Y, T, x, y, t, w, ϕ, and v, corresponding to the subject vectors and problem vectors defined above.

Thus, $X=[X_1 \ldots X_i \ldots X_N]$. It is a matrix of order $n \times N$ but, equivalently, it may be regarded as a supervector of order N with typical element Y_i. Similarly, the matrices Y, T, x, y, and t are all of order $n \times N$ and all may be regarded as supervectors of order N with typical elements Y_i, T_i, x_i, y_i, and t_i respectively. The matrix w is of order $p_1 \times N$ and may be regarded as a supervector with typical element w_i. The matrix ϕ is of order $p_2 \times N$ and may be regarded as a supervector of order N with typical element ϕ_i. Finally, the matrix v is of order $q \times n$ and may be regarded as a supervector of order n with typical element v_j.

These matrices are all schematized in Table 6, in which we summarize our notation.

Table 6. Summary of notation

n	problems
N	subjects
j, k	subscripts to index problems
l	subscript to index equivalent problems
i	subscript to index subjects
X_{ji}	$\begin{cases} 1, \text{ subject } i \text{ answers problem } j \text{ correctly} \\ 0, \text{ otherwise} \end{cases}$
Y_{ji}	$\begin{cases} 1, \text{ subject } i \text{ abandons problem } j \\ 0, \textit{ otherwise} \end{cases}$
T_{ji}	completion time for subject i on problem j
x_{ji}	realization of X_{ji}
y_{ji}	realization of Y_{ji}
t_{ji}	realization of T_{ji}
$P[X_{ji}=1 \mid Y_{ji}=1]=0$	
p_1	latent ability variables
p_2	latent continuance variables
$p=p_1+p_2$	latent trait variables
q	problem parameters
$X_i=(X_{1i}, X_{2i}, \ldots, X_{ji}, \ldots, X_{ni})^T$	
$Y_i=(Y_{1i}, Y_{2i}, \ldots, Y_{ji}, \ldots, Y_{ni})^T$	
$T_i=(T_{1i}, T_{2i}, \ldots, T_{ji}, \ldots, T_{ni})^T$	
$X=(X_1, X_2, \ldots, X_i, \ldots, X_N)$	
$Y=(Y_1, Y_2, \ldots, Y_i, \ldots, Y_N)$	
$T=(T_1, T_2, \ldots, T_i, \ldots, T_N)$	
$x_i=(x_{1i}, x_{2i}, \ldots, x_{ji}, \ldots, x_{ni})^T$	
$y_i=(y_{1i}, y_{2i}, \ldots, y_{ji}, \ldots, y_{ni})^T$	
$t_i=(t_{1i}, t_{2i}, \ldots, t_{ji}, \ldots, t_{ni})^T$	
$x=(x_1, x_2, \ldots, x_i, \ldots, x_N)$	
$y=(y_1, y_2, \ldots, y_i, \ldots, y_N)$	
$t=(t_1, t_2, \ldots, t_i, \ldots, t_N)$	
$w_i=(w_{1i}, w_{2i}, \ldots, w_{p_1 i})^T$	
$\phi_i=(\phi_{1i}, \phi_{2i}, \ldots, \phi_{p_2 i})^T$	
$v_j=(v_{j1}, v_{j2}, \ldots, v_{jq})^T$	
$w=(w_1, w_2, \ldots, w_i, \ldots, w_N)$	
$\phi=(\phi_1, \phi_2, \ldots, \phi_i, \ldots, \phi_N)$	
$v=(v_1, v_2, \ldots, v_j, \ldots, v_n)$	

We now consider the basic equations in our general model. They appear in Table 7.

Equation (1) states that if subject i has completion time T_{ji} on problem j, and does not abandon it, the (conditional) probability that he solves the problem correctly is some function $\alpha_{ji}(v_j, w_i, T_{ji})$ of the problem parameters v_j, of his own latent ability variables w_i, and of the completion time T_{ji}.

Equation (2) states that if subject i has completion time T_{ji} on problem j the (conditional) probability that he abandons the problem is a function of his own latent completion variables ϕ_i and of the completion time T_{ji}.

Equation (3) states that $\alpha_{ji}(v_j, w_i, T_{ji})$ is a cumulative logistic function $\psi[f_{ji}(\cdot)]$ of some function $f_{ji}(v_j, w_i, T_{ji})$ of the problem parameters v_j, of the subject variables w_i, and of the completion time T_{ji}.

Table 7. Equations for our general model

(1) $$P[X_{ji}=1|v_j, w_i, Y_{ji}=0, T_{ji}] \\ =\alpha_{ji}(v_j, w_i, T_{ji}) \\ =\alpha_{ji}$$

(2) $$P[Y_{ji}=1|\phi_i, T_{ji}] \\ =\beta_{ji}(\phi_i, T_{ji}) \\ =\beta_{ji}$$

(3) $$ln[\alpha_{ji}/(1-\alpha_{ji})]=f_{ji}(v_j, w_i, T_{ji})=f_{ji}=f$$

(4) $$ln[\beta_{ji}/(1-\beta_{ji})]=g_{ji}(\phi_i, T_{ji})=g_{ji}=g$$

(5) $$P[X_{ji}=0,\ Y_{ji}=0|v_j, w_i, \phi_i, T_{ji}] \\ =(1-\alpha_{ji})(1-\beta_{ji})$$

(6) $$P[X_{ji}=0,\ Y_{ji}=1|v_j, w_i, \phi_i, T_{ji}] \\ =\beta_{ji}$$

(7) $$P[X_{ji}=1,\ Y_{ji}=0|v_j, w_i, \phi_i, T_{ji}] \\ =\alpha_{ji}(1-\beta_{ji})$$

(8) $$P[X_{ji}=1,\ Y_{ji}=1|v_j, w_i, \phi_i, T_{ji}] \\ =0$$

(9) $$P[X_{ji}=x_{ji}|v_j, w_i, Y_{ji}=y_{ji}, T_{ji}] \\ =\alpha_{ji}^{x_{ji}}(1-\alpha_{ji})^{1-x_{ji}-y_{ji}}(1-x_{ji}y_{ji})$$

(10) $$P[Y_{ji}=y_{ji}|\phi_i, T_{ji}] \\ =\beta_{ji}^{y_{ji}}(1-\beta_{ji})^{1-y_{ji}}$$

(11) $$P[X_{ji}=x_{ji}, Y_{ji}=y_{ji}|v_j, w_i, \phi_i, T_{ji}] \\ =\alpha_{ji}^{x_{ji}}(1-\alpha_{ji})^{1-x_{ji}-y_{ji}} \\ \cdot\beta_{ji}^{y_{ji}}(1-\beta_{ji})^{1-y_{ji}}(1-x_{ji}y_{ji})$$

(12) $$L[v_j, w_i|X_{ji}=x_{ji},\ Y_{ji}=y_{ji}, T_{ji}=t_{ji}] \\ =\alpha_{ji}^{x_{ji}}(1-\alpha_{ji})^{1-x_{ji}-y_{ji}} \\ =L_{ji}[v_j, w_i]$$

(13) $$L[\phi_i|Y_{ji}=y_{ji},\ T_{ji}=t_{ji}] \\ =\beta_{ji}^{y_{ji}}(1-\beta_{ji})^{1-y_{ji}} \\ =L_{ji}[\phi_i]$$

(14) $$L[v_j, w_i, \phi_i|X_{ji}=x_{ji},\ Y_{ji}=y_{ji},\ T_{ji}=t_{ji}] \\ =\alpha_{ji}^{x_{ji}}(1-\alpha_{ji})^{1-x_{ji}-y_{ji}}\beta_{ji}^{y_{ji}}(1-\beta_{ji})^{1-y_{ji}} \\ =L_{ji}[v_j, w_i, \phi_i]=L_{ji}[v_j, w_i]\,L_{ji}[\phi_i]$$

(15) $$\frac{\partial\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial f_{ji}(v_j, w_i, t_{ji})}=[x_{ji}+(y_{ji}-1)\,\alpha_{ji}]$$

(16) $$\frac{\partial\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial g_{ji}(\phi_i, t_{ji})}=[y_{ji}-\beta_{ji}]$$

(17) $$\frac{\partial^2\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial f_{ji}(v_j, w_i, t_{ji})\,\partial g_{ji}(\phi_i, t_{ji})}=0$$

(18) $$E\left[\left(\frac{\partial\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial f_{ji}(v_j, w_i, t_{ji})}\right)^2\right] \\ =\alpha_{ji}(1-\alpha_{ji})(1-\beta_{ji})$$

(19) $$E\left[\left(\frac{\partial\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial g_{ji}(\phi_i, t_{ji})}\right)^2\right]=\beta_{ji}(1-\beta_{ji})$$

(20) $$E\left[\frac{\partial\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial f_{ji}(v_j, w_i, t_{ji})}\cdot\frac{\partial\, ln\, L_{ji}[v_j, w_i, \phi_i]}{\partial g_{ji}(\phi_i, t_{ji})}\right] \\ =0$$

(21) $$P[X_i=x_i|v, w_i,\ Y_i=y_i, T_i] \\ =\prod_{j=1}^{n} P[X_{ji}=x_{ji}|v_j, w_i,\ Y_{ji}=y_{ji}, T_{ji}]$$

(22) $$P[Y_i=y_i|\phi_i, T_i] \\ =\prod_{j=1}^{n} P[Y_{ji}=y_{ji}|\phi_i, T_{ji}]$$

(23) $$P[X_i, Y_i|v, w_i, \phi_i, T_i] \\ =\prod_{j=1}^{n} P[X_{ji}=x_{ji},\ Y_{ji}=y_{ji}|v_j, w_i, \phi_i, T_{ji}]$$

(24) $$L[v, w_i, \phi_i|X_i=x_i,\ Y_i=y_i,\ T_i=t_i] \\ =\prod_{j=1}^{n} L_{ji}[v_j, w_i, \phi_i] \\ =L_i[v, w_i, \phi_i]$$

(25) $$L[v, w_i|X_i=x_i,\ Y_i=y_i,\ T_i=t_i] \\ =\prod_{j=1}^{n} L_{ji}[v_j, w_i] \\ =L_i[v, w_i]$$

(26) $$L[\phi_i|Y_i=y_i,\ T_i=t_i] \\ =\prod_{j=1}^{n} L_{ji}[\phi_i] \\ =L_i[\phi_i]$$

(27) $$L[v, w, \phi|X=x,\ Y=y,\ T=t] \\ =\prod_{i=1}^{N} L_i[v, w_i, \phi_i] \\ =L[v, w, \phi]$$

(28) $$L[v, w|X=x,\ Y=y,\ T=t] \\ =\prod_{i=1}^{N} L_i[v, w_i] \\ =L[v, w]$$

(29) $$L[\phi|Y=y,\ T=t] \\ =\prod_{i=1}^{N} L_i[\phi_i] \\ =L[\phi]$$

(30) $$L[v, w, \phi]=L[v, w]\,L[\phi]$$

(31) $$ln\, L[v, w, \phi]=ln\, L[v, w]+ln\, L[\phi]$$

(32) $$ln\, L[v, w]=\sum_{i=1}^{N} ln\, L_i[v, w_i]$$

(33) $$ln\, L[\phi]=\sum_{i=1}^{N} ln\, L_i[\phi_i]$$

(34) $$L[v, w]=L[v|w] \\ =L[w|v]$$

(35) $$L_i[v, w_i]=L_i[w_i|v]$$

Equation (4) states that $\beta_{ji}(\phi_i, T_{ji})$ is a cumulative logistic function $\psi[g_{ji}(\cdot)]$ of some function $g_{ji}(\phi_i, T_{ji})$ of the subject variables ϕ_i, and of the completion time T_{ji}.

Given Eqs. (1) and (2), the law of compound probabilities, and our definitions of the random variables X_{ji} and Y_{ji}, Eqs. (5)–(8) follow directly. These, of course, are the formulae for the four logical response patterns on the two dichotomous variables X_{ji} and Y_{ji}.

This completes the statement of the general model in that we now have expressions in α_{ji} and β_{ji} for all response patterns, contingent on completion at time T_{ji}. But, of course, α_{ji} and β_{ji} are functions of $f_{ji}(v_j, w_i, T_{ji})$ and of $g_{ji}(\phi_i, T_{ji})$ and these functions are still unspecified. Indeed, at this point even the problem parameters v_j and the latent trait variables w_i and ϕ_i are still unspecified.

And, of course, we are still working at the level of the single subject and the single problem. Before introducing more complexities we present some results which we will need subsequently.

Equation (5) is the probability, conditional upon completion at time T_{ji}, that subject i will yield an error response to problem j.

Equation (6) is the probability, conditional upon completion at time T_{ji}, that subject i will abandon problem j.

Equation (7) is the probability, conditional upon completion at time T_{ji}, that subject i will yield a correct response to problem j.

Equation (8) states only that regardless of completion time T_{ji}, regardless of the problem parameters v_j, and regardless of the subject parameters (w_i and ϕ_i) if subject i abandons problem j he cannot yield a correct solution to problem j.

Given Eqs. (1) and (2) and Eqs. (5)–(8), expressions for the followings six probabilities follow directly:

(i) P [correct | not abandon, completion at time T_{ji}]
(ii) P [incorrect | not abandon, completion at time T_{ji}]
(iii) P [correct | abandon, completion at time T_{ji}]
(iv) P [incorrect | abandon, completion at time T_{ji}]
(v) P [abandon | completion at time T_{ji}]
(vi) P [not abandon | completion at time T_{ji}]

Equation (9) expresses (i)–(iv) simultaneously and Eq. (10) expresses (v) and (vi) simultaneously. It is easy to verify that Eq. (11) expresses Eqs. (5)–(8) simultaneously. We note that Eq. (11) is the product of Eq. (9) and Eq. (10).

Equation (11) allows us to evaluate the probability of a particular response pattern for hypothetical values of the problem parameters v_j, of the subject's latent trait variables, w_i and ϕ_i, and of the completion times T_{ji}. Once the subject responds, however, we have observed values for x_{ji}, y_{ji}, and for t_{ji} and we may regard Eq. (11) as a function of the unobservable $\cdot v_j$, w_i, and ϕ_i conditional on the observed data x_{ji}, y_{ji}, and t_{ji}. This function is called a likelihood function. Similar reasoning leads to likelihood functions corresponding to Eqs. (9) and (10). The likelihood functions corresponding to Eqs. (9)–(11) appear in Eqs. (12)–(14) respectively. They provide the building blocks with which we proceed. Each is equal numerically to the corresponding probability formula and each is provided with an abbreviated form for use in subsequent expressions. Equations (9) and (11) contain the term $(1-x_{ji}y_{ji})$. This is a device which provides for the impossible case $x_{ji}=1$, $y_{ji}=1$. Given valid data, however, this case does not occur and the term $(1-x_{ji}y_{ji})$ is deleted from the corresponding likelihood functions.

We noted above that Eq. (11) is the product of Eq. (9) and (10). This, of course, is a special case of the law of compound probabilities. In the same way, Eq. (14) is the product of Eq. (12) and (13). The important thing to note here is that one factor [(Eq. (12)] involves α_{ji}, but not β_{ji}, and the other factor [(Eq. (13)] involves β_{ji} but not α_{ji}. This property of separability, as it has been called, is displayed in two other forms in Eqs. (15)–(17) and in Eqs. (18)–(20).

It is usual, when working with models involving likelihood functions, to work with the logarithm of the likelihood function which is monotonic with the likelihood function. This leads to additive factors rather than to multiplicative factors and the resulting expressions are frequently much simpler in form. The present case is no exception.

Equations (15)–(17) display the partial derivatives of $\ln L_{ji}[v_j, w_i, \phi_i]$ with respect to $f_{ji}[v_j, w_i, t_i]$ and $g_{ji}[\phi_i, t_{ji}]$.

If the response is an abandonment then y_{ji} is 1 and x_{ji} is 0 and Eq. (14) vanishes. Otherwise, Eq. (1) reduces to $x_{ji}-\alpha_{ji}$, which has the same form as Eq. (16). The separability property is demonstrated dramatically in Eq. (17).

The separability property is illustrated in a different form in Eqs. (18)–(20). These are the expressions for the elements in Fisher's information matrix. The separability property is once more illustrated dramatically – this time in Eq. (20). We exploit this property fully in the following development. It means in fact that our model factors into two submodels. One submodel is concerned with abandonment versus non-abandonment and the other submodel is concerned with success versus failure given non-abandonment.

This completes, for the moment, our development of the general model at the level of the individual subject and the single problem. We now consider what happens when the individual subject is confronted with a set of n problems.

We state the assumption of local independence for our general model in Eqs. (21)–(23). Evidently, any two of these three equations imply the third.

Equation (23) looks more like those given for the more simple latent trait models which have been proposed in the literature. But it, alone, is insufficient for our purposes. Since eqs. (21) and (22) are more simple in form than is Eq. (23) we take them as our definition of local independence and regard Eq. (23) as a consequence of this definition. Both Eqs. (21) and (22) specify that the conditional responses thus defined are statistically independent. We consider each equation in turn and then discuss the implications of local independence in more detail.

Equation (21) defines the probability of a particular pass-fail response pattern on the set of n problems conditional on the abandon-non-abandon response pattern on the set of problems, conditional on the problem parameters v, conditional on the subjects latent ability parameters w_i and conditional on the completion times T_i on the set of problems.

Similarly Eq. (22) defines the probability of a particular abandon-non-abandon response pattern on the set of n problems conditional on the subjects latent continuance variables ϕ_i and conditional on the completion times T_i on the set of problems.

Equation (23) states the probability of the pass-fail-abandon-non-abandon response pattern on the set of n problems conditional on the problem parameters v, conditional on the subjects latent ability variables w_i, conditional on the subjects latent continuance variables ϕ_i, and conditional on the completion times T_i on the set of problems.

We now consider a population of hypothetical subjects each of whom has the same latent ability variables and the same latent continuance variables and each of whom has the same values on all latent trait variables which contribute to completion time. In such a population all subjects will have identical expected completion times on a set of n problems.

Local independence implies that for such a population, homogeneous in all relevant respects, response patterns to different problems are statistically independent. It does not, however, imply that such response patterns are independent in a population heterogeneous with respect to the latent trait variables. Indeed, the converse is the rule rather than the exception. When we find a relationship between such response patterns in a heterogeneous population and note that the relationship disappears within homogeneous subpopulations we say that we have accounted for the relationship and we regard the latent trait variables as the determinants of the relationship.

In Eqs. (24)–(26) we write the likelihood functions corresponding to Eqs. (21)–(23). Equation (24) gives the likelihood of the parameters. v, of the problems and of the latent ability variables, w_i, and latent continuance variables, ϕ_i, of the subject conditional on the data x_i, y_i, and t_i. Equation (25) gives the likelihood of the parameters, v, of the problems, and of the latent ability variables, w_i, of the subject conditional on the data x_i, y_i, and t_i. And Eq. (26) gives the likelihood of the latent continuance variables, ϕ_i, of the subject conditional on the data y_i and t_i.

We note that Eq. (23) is the product of Eqs. (21) and (22) and that Eq. (24) is the product of Eqs. (25) and (26). Here, once more, we see the same pattern that we have seen already in Eqs. (11) and (14) – the law of compound probabilities. This time, though, we have paid for it with the assumptions made in Eqs. (21) and (22), which gave us local independence.

We now assume statistical independence across subjects and build up the likelihood for the problem parameters, v, and the subject latent trait variables w and ϕ conditional on the entire data set x, y, and t. Equations (27)–(29) are the products, across subjects, of Eqs. (24)–(26) respectively. They follow directly from the condition for statistical independence.

Equation (29) forms the basis for all of what follows on the general model. It is numerically equivalent to the probability of the data set $X=x$, $Y=y$ conditional on the problem parameters, v, on the latent trait variables, w and ϕ, and on the completion times t. In Eq. (29), however, we regard it as a function of v, w, and ϕ conditional on the observed data x, y, and t.

This leads naturally to the method of maximum likelihood estimation which seeks numerical values for the problem parameters, v, and for the latent trait variables w and ϕ, which maximize the value of Eq. (29) conditional on the observed data x, y, and t. It thus seeks for problem parameters and latent trait variables which, given the general model, are most consistent, in the sense just described, with the observed data.

Before we discuss some ways in which we might utilize the maximum likelihood principle in our estimation problem we require some more notation.

Equation (30) states formally that Eq. (29) is the product of Eqs. (27) and (28) and Eq. (31) states that the natural logarithm of Eq. (29) is the sum of the natural logarithms of Eq. (27) and (28). Since Eqs. (27) and (28) are each products of likelihoods their respective logarithms are the corresponding sums of logarithms noted in Eqs. (32) and (33).

In Eq. (34) we note that the likelihood expressed in Eq. (29) may be regarded as a function of v and w jointly, as a function or v conditional on a given w, or as a function of w conditional on a given v. Similarly, in Eq. (35), we note that Eq. (25) may be regarded as a function of v and w_i jointly or as a function of w_i conditional on a given v. Equations (31)–(33) provide the basis for two potentially powerful algorithms for maximizing Eq. (29). Equations (34) and (35) are purely notational but their use simplifies greatly our description of the algorithms.

As we noted earlier, the likelihood function expressed in Eq. (27) is numerically equal to a probability function which is the probability of the data conditional on the unobservable problem parameters, the unobservable latent ability variables, the unobservable latent continuance variables, and the response times or latencies. Thus, our probabilistic model leads quite naturally to a maximum likelihood method for the estimation of the unobservable problem parameters and latent trait variables in the model.

The likelihood function defined in Eq. (27) involves qn unknown problem parameters, $p_1 N$ unobservable latent ability variables, and $p_2 N$ latent continuance variables.

To maximize a general function involving so many unknowns is a very formidable task indeed. We capitalize on some special properties of our general model and reduce this task to manageable proportions.

We note [(Eq. (30)] that the likelihood function factors into two parts. The first part involves the problem parameters, v, and the latent ability parameters w. The second part involves the latent continuance variables, ϕ. The logarithm of the likelihood function is thus the sum of two parts – one involving the problem parameters and the latent ability parameters and the other involving the latent continuance variables. Since both factors in Eq. (30) are probability functions, both terms in Eq. 31 are negative. Since each unknown is involved in only one of the two terms it follows that to maximize Eq. (30) it is sufficient to maximize each term in Eq. (31) separately. We thus split the general maximization problem involving $qn+(p_1+p_2)N$ unknowns into two independent subproblems involving $qn+p_1N$ and p_2N unknowns respectively. This implies a drastic reduction in effort relative to the general problem.

We now consider the second term in Eq. 31 and its representation in Eq. (33). Since each factor in Eq. (29) is a probability function, each term in Eq. (33) is negative. Since each term in Eq. (33) involves one, and only one, of the unknown latent ability variables it follows that to maximize the second term in Eq. (31) it is sufficient to maximize each term in Eq. (33) separately. We thus split the subproblem involving p_2, N unknowns into N independent subproblems, each involving p_2 unknowns. This implies a further drastic reduction in effort.

When we turn to the first term in Eq. (31) things are not quite so simple since each term in Eq. (32) involves the unknown problem parameters. However, as we indicated in Eq. (34), we may regard $L[v, w]$ as a function of v and w jointly, as a function of v conditional on given values for w, or as a function of w conditional on given values for v. Similarly [Eq. (35)], we may regard $L_i[v, w_i]$ as a function of v and w_i jointly, as a function of v conditional on given values for w_i, or as function of w_i conditional on given values for v. Looked at this way the problem of maximizing $L[v, w]$, conditional on given values for v, splits up [Eq. 32] into N independent conditional maximization subproblems, each involving its own w_i. We exploit this fact in two iterative algorithms for maximizing the first term in Eq. (31). We now describe these two algorithms. (Algorithm I is sketched in Table 8).

Table 8. Algorithm I

Start with v^0, w^0.

Find w^1 such that
$L[v^0, w^1] = \max_w L[v^0, w]$.

Then find v^1 such that
$L[v^1, w^1] = \max_v L[v, w^1]$.

In general, given v^k, w^k, and hence $L[v^k, w^k]$, find w^{k+1} such that
$L[v^k, w^{k+1}] = \max_w L[v^k, w]$.

Then find v^{k+1} such that
$L[v^{k+1}, w^{k+1}] = \max_v L[v, w^{k+1}]$.

Continue until

$$\begin{aligned} L[v^{k+1}, w^{k+1}] &= \max_w L[v^{k+1}, w] \\ &= \max_v L[v, w^{k+1}] \\ &= \max_v \max_w L[v, w]. \end{aligned}$$

Algorithm I

Each iteration consists of two stages. We fix the problem parameters at their current values and update the latent ability variables. We then fix the latent ability variables at their new values and update the problem parameters. We proceed in this way until the relative change in the problem parameters is below some specified small value.

Similar two-stage algorithms have been described by Birnbaum (1968), Lord (1968), Wright and Douglas (1972), Fischer (1974), and Andrich (1978).

The algorithm which we have described above (algorithm 1) differs from the others in that we maximize the likelihood function directly, whereas the other procedures seek for solutions to the sets of simultaneous non-linear equations which result when the partial derivatives of the logarithm of the likelihood with respect to the latent ability variables and with respect to the problem parameters are equated, respectively, to zero.

Our main reason for chossing a two-stage iterative procedure is that a massive problem is thus split up into a very large number of much smaller subproblems, each of which is quite manageable. Our reason for choosing to maximize likelihood functions is that we avoid one of the main problems associated with the latter procedures. This problem is that with some systems of equations the iterations diverse rather than converse unless the initial estimates of the parameters are very close to the optimum values. By maximizing the likelihood function directly we never accept a trial solution unless it increases the goodness of fit.

Two other major problems with the likelihood equations approach is that not infrequently the values for successive iterates tend to oscillate about the optimal values or if they converge they tend to converge very slowly indeed. Unfortunately, algorithm I as described above is prone as well to both of these latter problems. We programmed an algorithm due to Ramsay (1975) and incorporated it into our two-stage algorithm and the resultant procedure seemed to cope effectively with both problems. Now we turn to a description of our algorithm II. (Algorithm II is sketched in Table 9).

Table 9. Algorithm II

$$L[v, w^*] = \max_w L[w|v]$$

$$L[v^*, w] = \max_v L[v|w]$$

$$\begin{aligned} L[v^*, w^*] &= \max_v [\max_w L[w|v]] \\ &= \max_w [\max_v L[v|w]] \\ &= \max_v \max_w L[v, w] \end{aligned}$$

Algorithm II

This is a nested procedure. An outer loop iterates on the problem parameters v and, within this loop, an inner loop iterates on the subject's latent trait variates w. As in step 2 of algorithm I the inner optimization is achieved on a subject-by-subject basis and, as before, a very large optimization problem is reduced to a sequence of much smaller subproblems.

This procedure does not suffer from the drawbacks, noted above, associated with algorithm I. It does, though, depend critically on the optimization in the inner loop being computed very quickly. Our first effort to apply the algorithm failed miserably. We were never able to complete even a single iteration in the outer loop because we always ran out of computer time first. It was only when we introduced the regression hypothesis (i.e that 'speed' is the regression of problem difficulty on solution time) that algorithm II became of more than theoretical significance. The reason for this is that the regression hypothesis implies that for a particular subject the probability of a correct response given non-abandonment and completion at time is constant across problems, and this, as we shall see, leads to closed-form expressions for a subject's speed and accuracy.

Estimation for Individual Subjects

We now consider in some detail the estimation of speed and accuracy for an individual subject under the regression hypothesis. Let us assume that a particular subject, i, with mean time to non-abandonment, $\bar{t}_{\mathrm{RW}}$, on a set or n problems achieves R correct responses, W wrong, or incorrect, responses, and A abandonments.

A non-abandonment (R or W) implies that

$$\ln \frac{\alpha}{1-\alpha} = a + S(t - \bar{t}_{\mathrm{RW}}) - d$$

or, equivalently, that

$$\ln \frac{\alpha}{1-\alpha}+d=a+S(t-\bar{t}_{\mathrm{RW}}).$$

An abandonment, on the other hand, implies that $\ln \frac{\beta}{1-\beta}=t-p$, or, equivalently, that

$$\ln \frac{1-\beta}{\beta}+t=p.$$

Evidently, we have n equations involving $\ln \frac{\alpha}{1-\alpha}$ and $\ln \frac{1-\beta}{\beta}$ as linear functions of speed (s), accuracy (a), and persistence (p).

Under the regression hypothesis,

$$\{E[S(t-\bar{t}_{\mathrm{RW}})]=d \mid \text{non-abandonment}\},$$

it is evident that the probability of a correct response given non-abandonment is constant on the set of R+W non-abandoned problems. It is also evident that, under the constancy hypothesis, the probability of an abandonment is constant on the set of A abandoned problems.

We thus have R+W independent observations from a binomial (R+W, α) distribution $\alpha^{\mathrm{R}}(1-\alpha)^{\mathrm{W}}$, and A independent observations from a binomial (A,β) distribution $\beta^{\mathrm{A}}(1-\beta)^{\mathrm{R+W}}$.

It follows from standard and well-known results that $\hat{\alpha}=\frac{\mathrm{R}}{\mathrm{R+W}}$ is the maximum likelihood estimate of α and that $\hat{\beta}=\frac{\mathrm{A}}{\mathrm{R+W+A}}$ is the maximum likelihood estimate of β. If R or W is zero, then $\hat{\alpha}$ does not exist and if A or R+W is zero, then $\hat{\beta}$ does not exist.

The tableau in Table 10 displays the equations for a hypothetical subject who has correct responses on probelms 1, 4, 7, 10 and 11, wrong responses on problems 2, 5, 9, and 12, and abandonments on problems 3, 6, and 8. We have, then, $n=12$ problems, R=6 correct responses, W=4 incorrect or wrong responses, and A=3 abandonments. We have nine independent responses from a binomial (9, α) distribution, $\alpha^5(1-\alpha)^4$ and three independent responses

Table 10. Estimation for individual subjects

$$\underbrace{\begin{bmatrix} ln\frac{\alpha_i}{1-\alpha_i}+d_1 \\ ln\frac{\alpha_i}{1-\alpha_i}+d_2 \\ ln\frac{1-\beta_i}{\beta_i}+t_{3i} \\ ln\frac{\alpha_i}{1-\alpha_i}+d_4 \\ ln\frac{\alpha_i}{1-\alpha_i}+d_5 \\ ln\frac{1-\beta_i}{\beta_i}+t_{6i} \\ ln\frac{\alpha_i}{1-\alpha_i}+d_7 \\ ln\frac{1-\beta_i}{\beta_i}+t_{8i} \\ ln\frac{\alpha_i}{1-\alpha_i}+d_9 \\ ln\frac{\alpha_i}{1-\alpha_i}+d_{10} \\ ln\frac{\alpha_i}{1-\alpha_i}+d_{11} \\ ln\frac{\alpha_i}{1-\alpha_i}+d_{12} \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 1 & t_{1i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 1 & t_{2i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 0 & 0 & 1 \\ 1 & t_{4i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 1 & t_{5i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 0 & 0 & 1 \\ 1 & t_{7i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 0 & 0 & 1 \\ 1 & t_{9i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 1 & t_{10i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 1 & t_{11i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \\ 1 & t_{12i}-\bar{t}^{(i)}_{\mathrm{RW}} & 0 \end{bmatrix}}_{Y} \underbrace{\begin{bmatrix} a_i \\ s_i \\ p_i \end{bmatrix}}_{\theta}$$

from a binomial (3, β) distribution, $\beta^3(1-\beta)^9$. We have a system of 12 linear equations relating the unknown latent trait variates a_i, s_i, and β_i to α_i and β_i, which are both unknown, and to the problem difficulties for the nine non-abandoned problems which, of course, are also unknown.

In general, for $i=1,2,\ldots,N$ subjects we have

$$\mathrm{R}_i=\sum_{j=1}^{n} x_{ji}, \qquad \mathrm{A}_i=\sum_{j=1}^{n} y_{ji},$$

and

$$\mathrm{W}_i=n-\sum_{j=1}^{n} x_{ji}-\sum_{j=1}^{n} y_{ji}$$

(number right, number abandoned, and number wrong respectively) and for each subject we have for $j=1,2,\ldots,n$ problems

n equations of the form

$$y_{ji}\left[ln\frac{1-\beta_i}{\beta_i}+(t_{ji}-p_i)\right]$$
$$=(y_{ji}-1)\left[ln\frac{\alpha_i}{1-\alpha_i}+a_i+s_i(t_{ji}-\bar{t}^{(i)}_{RW})-d\right].$$

It is quite clear for the example in Table 10 that column 3 of the matrix X is orthogonal to both columns 1 and 2 and, since the sum of deviations from a mean are zero, column 2 is orthogonal to column 1. It is evident that this is true in general and thus that in general $X'X$ is diagonal.

A bit of simple algebra shows that the normal equations $X'X\theta=X'y$ have a very simple form and it follows that $\theta=(a_i, s_i, p_i)'$ has the elements shown below. In these equations $\bar{t}_{RW}$, $\bar{t}_A$, and α_{RW} are, respectively, mean time to non-abandonment, mean time to abandonment, and mean difficulty of non-abandoned problems for subject i.

If $\hat{\alpha}_i$ is the maximum likelihood estimate of α_i it follows that, since $\ln\frac{\hat{\alpha}_i}{1-\hat{\alpha}_i}$ is a one-to-one function of α_i, $\ln\frac{\alpha_i}{1-\hat{\alpha}_i}$ is the maximum likelihood estimate of $\ln\frac{\alpha_i}{1-\alpha_i}$. It follows that, since $\hat{\alpha}_i=\frac{R_i}{R_i+W_i}$, $\ln\frac{R_i}{R_i+W_i}$ is the maximum likelihood estimate of $\ln\frac{\alpha_i}{1-\alpha_i}$. Similarly, if follows that $\ln\frac{R_i+W_i}{A_i}$ is the maximum likelihood estimate or $\ln\frac{1-\beta_i}{\beta_i}$. Substituting these m.l.e. estimates for $\ln\frac{\alpha_i}{1-\alpha_i}$ and $\ln\frac{1-\beta_i}{\beta_i}$, respectively, yields the following estimates for speed, accuracy, and persistence:

$$\hat{s}_i=\frac{\sum_j d_j(t_{ji}-\bar{t}^{(i)}_{RW})(1-y_{ji})}{(t_{ji}-\bar{t}^{(i)}_{RW})^2(1-y_{ji})},$$

$$\hat{a}_i=ln\frac{R_i}{W_i}+\bar{d}^{(i)}_{RW},$$

and

$$\hat{p}_i\doteq ln\frac{R_i+W_i}{A_i}+\bar{t}^{(i)}_A.$$

Table 11. Some information matrices

$$I[a_i^*, s_i^*]$$
$$=\sum_{j=1}^{n}\alpha_{ji}(1-\alpha_{ji})(1-\beta_{ji})D_j^2\begin{bmatrix}1 & T_{ji}\\ T_{ji} & T_{ji}^2\end{bmatrix}$$
$$I[a_i, s_i]$$
$$=\sum_{j=1}^{n}\alpha_{ji}(1-\alpha_{ji})(1-\beta_{ji})D_j^2\begin{bmatrix}1 & (T_{ji}-\bar{T}_i)\\ (T_{ji}-\bar{T}_i) & (T_{ji}-\bar{T}_i)^2\end{bmatrix}$$

These formulae include the unknown problem parameters d_j. At each iteration we replace them with their current best estimates. At convergence, of course, these too are maximum likelihood estimates. The final, speed, accuracy, and persistence estimates are thus known and quite simple one-to-one functions of the maximum likelihood estimates for $\ln\frac{\alpha_i}{1-\alpha_i}$, $\ln\frac{1-\beta_i}{\beta_i}$, and or the d_j. Thus, the estimates of speed, accuracy, and persistence are maximum likelihood estimates.

In Table 11 we display information matrices for two versions of the speed- accuracy submodel.

First we consider the form

$$\ln\frac{\alpha_i}{1-\alpha_i}=a_i+s_i\,t_{ji}-d$$

in which response times to non-abandonment are in raw form. Since these times are all non-negative, their sum must be non-negative and, in general, is positive. It is easy to show, then, that the inverse of this matrix to which the variance-covariance matrix of the estimates $\hat{s}_i$ and $\hat{a}_i$ is proportional must have negative off-diagonal elements. In the reparameterized form

$$\ln\frac{\alpha_i}{1-\alpha_i}=a_i+s_i(t_{ji}-\bar{d}^{(i)}_{RW})-d_j,$$

in which the times to non-abandonment are in deviation units their sum is zero and the information matrix and the variance-covariance matrix for the estimates are both diagonal.

In the first form if speed is overestimated then accuracy is underestimated and vice versa. In the second form this dependency is eliminated. In reparameterizing the model in this way we have not done anything mysterious. Indeed the estimates for the second form are identical to those yielded by the traditional formulae for linear regression, which may be found in any elementary statistics book and which are programmed in most of the more recent scientific pocket calculators.

We chose the maximum likelihood method for parameter estimation because it seemed so naturally to fit with the probabilistic formulation of our models. After we had devised a maximum likelihood algorithm which seemed to be both effective and practicable, we decided to investigate a weighted least squares procedure with which we had become familiar in fitting models from biometrical genetics to sets of empirical data.

We outline the method in terms of fitting the speed-accuracy submodel to the pass-fail data conditional on non-abandonment (i.e. $X=1$ or 0 for correct and incorrect, respectively, given that $y=1$ for abandonment and $y=0$ for non-abandonment).

The loss function for a single observation is $G=[x-E(x)]^2/\mathrm{Var}[E(x)]$. For our model $E(x)$, given non-abandonment, is α and $V[E(x)]$ is just $\alpha(1-\alpha)$. Thus, it follows directly that $G=(x-\alpha)^2/\alpha(1-\alpha)$ and that $G=(1-\alpha)/\alpha$ or $\alpha/(1-\alpha)$ as x is, respectively, 1 or 0. In our case we have

$$\ln\frac{\alpha}{1-\alpha}=a+s(t-\bar{t}_{\mathrm{RW}})-d=f,\quad \text{say.}$$

It follows that, if $x=1$, $G=\exp(-f)$ and, if $x=0$, $G=\exp(f)$. Since least squares loss functions seem to be nearly quadratic near the minimum for well-behaved functions and since most minimization procedures have been designed to have optimum convergence properties for quadratic functions, it seemed natural to try to capitalize on the obvious computational simplicity.

Under the hypothesis of local independence the responses for a single subject are independent and under the natural assumption that the responses of different subjects are independent the loss function for the data set is

$$G=\sum_j\sum_i[x_{ji}-E(x_{ji})]^2/V[E(x_{ji})].$$

If we had replaced the denominator by unity, then, of course, our loss function would have been ordinary (unweighted) least squares. Thus we may regard each term in G as a squared residual $(x-E(x)]^2$ multiplied by the weighting factor $1/V[E(x)]$. This weighting factor, the reciprocal of the variance, is known as the precision of the estimate of the observation. Thus the squared residual is given a weight which is proportional to the precision of the estimate of the observation.

In most applications of weighted least squares the function has the form

$$G=[S-E(S)]^2/V[E(S)],$$

where S is a statistic which is based on a large sample, $E(S)$ is its expectation under the model, and $V[E(S)]$ is the variance of this estimate of the statistic. In such cases the residuals have been shown to be distributed approximately as chi-squared. In the present case, however, such details are quite irrelevant since our 'statistics' are but single observations.

An interesting aspect of G in the current application is that $\partial G/\partial f$ is $-\exp(-f)$ if $x=1$ and is $\exp(f)$ if $x=0$. Thus the cost of computing first-order derivatives is negligible once we compute G. A curious property is that the second derivative $\partial^2 G/\partial f^2=G$, and thus a method using second-order partial derivatives requires little additional work once we have G and the first-order partial derivatives. We have not capitalized on this remarkable property since we have concentrated on quasi-Newton methods, which do not require the user to program explicit formulae for the matrix of second-order partial derivatives.

The modification of our program to incorporate a weighted least squares procedure rather than maximum likelihood took no more than a few minutes to implement.

The result was quite dramatic. The iterations seemed to proceed with fewer problems and computational time was reduced to well less than half that required by the maximum likelihood procedure.

Goodness of Fit and Model Assessment

One problem which the serious worker in this area will inevitably encounter is that of comparing the goodness of fit of one model with that of another. This is not a simple matter. We know of no completely general statistical methods for this purpose. There are two important special cases, however, for which standard methods are available.

The first case is that in which both models have the same number of parameters and the parameters may be estimated by the same method. In this case the distribution of residuals under the two models may be compared to determine which model shows the better fit.

The second case is that in which a more complex model is constructed by adding parameters to another, more simple, model. In this case the simpler model is said to be nested within the more complex model. Under these circumstances it is sometimes possible that the improvement in goodness of fit due to the additional parameters may be tested for statistical significance.

Hanna (1969), takes an information-theoretic approach to the problem. He points out that the traditional methods, to which we have just alluded, treat parameters as 'degrees of freedom'. He argues, forcibly, that this view may, in some circumstances at least, be downright misleading and that, in general, in comparing models, the number of parameters is largely irrelevant.

He derives a number of information measures which relate to the information content of a model before and after estimation of the parameters. Together, he says, they provide a measure of the information taken from the data by the estimation process. He goes on to differentiate between the descriptive power of a model and its predictive (or explanatory) power. He then considers the notion of parameter invariance of a model and that of the falsifiability of a model. He also differentiates between the extent to which a model takes information from the environment (the properties of the problems as selected by the tester) and the extent to which it takes information from the data (from the responses of the subject to the particular problems selected). He not only makes these distinctions; he also quantifies each notion in a form which permits comparison between and among models independently of the numbers of parameters in the respective models. A consequence of his information-theoretic approach to model testing is that a model which is statistically rejected in terms of goodness of fit may be regarded as more effective empirically than the statistically accepted models. He argues that a model which cannot be statistically rejected on the basis of given evidence may provide less information about the empirical variables influencing behaviour than an alternative model which can be rejected statistically and that the more informative model is preferable regardless of whether or not it can be statistically rejected.

Although we have not, in the decade since its appearance, seen a single citation of his paper, we regard it as the most important paper on its topic that we have yet seen.

In terms of our own work such issues seem at the moment to be of mainly academic interest. We are still trying to build a model which takes into account some of the more important factors involved in problem solving behaviour. We have not yet fitted a simple unidimensional model to our data. Indeed, we feel that such an endeavour would be but a sterile statistical exercise even if we could find a competing model which addresses itself to the same data. In fact, no such competing model exists.

An Application of the Model

We now present an application of our model to a fairly extensive set of empirical data which we collected in collaboration with Professor Eysenck. Our main thrust here, since we view this investigation as a methodological investigation rather than as a substantive one, is to illustrate, convincingly, we think, that non-trivial models of the sort we have proposed above may in practice be applied and that results derived therefrom may, in some cases at least, be readily interpreted.

In the testing situation groups of up to four children were tested simultaneously and data regarding pass-fail, abandon-not-abandon and response latency were collected for each child on every problem. The subject sits before a testing console and the apparatus presents a multiple-choice question on a display screen using back projection from an automatic slide projector. Push buttons are provided for the subject's responses, including one to indicate that the subject does not know the answer and wishes to proceed to the next problem. When a button is pressed, the slide projector is stepped on to present the next problem. During the time that the slide is being changed a voltage output is produced by the apparatus, the voltage level being related to the button that was pressed. At all other times the voltage level is zero. Thus, the appearance of a non-zero voltage at the output of the apparatus defines the time at which the subject makes a response, the disappearance of this voltage defines the time at which the next problem is presented, and the level of the voltage defines the particular response made. The output of the apparatus feeds into one of four channels of an analog instrumentation tape recorder. Thus, with four testing consoles, up to four subjects may be tested simultaneously and complete information regarding all responses (including abandonments) and their latencies may be retrieved from the tape. The tape thus produced is processed on the Linc-8 computer system where the analog information is digitalized, the responses and latencies are decoded and verified, and the results are punched onto paper tape for subsequent statistical analysis.

Table 12. Composition of sample

Boys	109
Girls	85
Total	194
Age range	14–16 years
Mean age	15·082 years
Std. deviation	0·399 years

For the moment it suits our purpose to note only that the battery consists of four subtests which were labelled, respectively, Anagrams, Mill Hill, Numbers, and Ravens. Clearly, they attempted to select problems which span a range of cognitive abilities.

In Table 12 we give the composition of our sample in terms of sex and age. We will see, later, that performance on the cognitive tests, as summarized by our estimates of speed, accuracy, and persistence, seems to be independent of these factors.

In the top part of Table 13 we display, for each subtest, the number of subjects for whom persistence scores exist and the number of subjects for whom both speed and accuracy scores exist. In the bottom part of Table 13 we display for each combination of speed, accuracy, and persistence, including self combinations, the number of subjects for whom scores on all such measures exist. Thus, there are 127 subjects for whom speed estimates, and thus accuracy estimates as well, exist on all four subtests.

Table 13. Sizes of matrices

	Speed and accuracy		Persistence
Anagrams	171		158
Mill Hill	183		135
Numbers	187		141
Raven	159		36
	Speed	Accuracy	Persistence
Speed	127	127	76
Accuracy	–	127	76
Persistence	–	–	98

There are only 86 subjects for whom persistence measures exist on all three subtests remaining after we exclude the Ravens subtest. We exclude the latter subtest because, as may be seen in the top part of Table 13, there are only 36 subjects for whom such scores exist. Furthermore, there are only 76 subjects for whom both speed and persistence scores, and thus both accuracy and persistence scores, exist.

Clearly, to restrict our analysis to all subjects for whom speed, accuracy, and persistence scores exist on all subtests would ignore most of our data. Although such an analysis, with complete data for all subjects, would be most straightforward to implement, the cost of such simplicity seems to be prohibitive.

On the other hand, the analyses which we wish to carry out on these data have not, at the present time at least, been formulated for the general case with arbitrary missing data. Nor do we see such analyses being viable in the forseeable future. Consequently, in order to pursue our aims we see no alternative to compromise. The analyses we now describe seem to strike a reasonable balance between the readily attainable, but trivial, extreme on the one hand and the ideal, but unattainable, extreme on the other.

We now work, step by step, through the analyses which we have carried out.

1. The model was fitted separately to problems from each subtest. The weighted least squares procedure, which we have described above, was employed. This step resulted in a set of problem difficulties for each subtest. These we display in the top part of Table 14. Although these values are on a quite arbitary, and unknown, scale there seem to be remarkable similarities among the four sets in terms of scale and origin. This impression is confirmed, and possibly strengthened, by the regression equations displayed in the bottom part of Table 14. Here, we computed a separate linear regression for each subtest. The dependant variable in each case is the proportion of subjects who fail the problem conditional on the event that the problem was not abandoned. These values, by the way, were chosen as starting values to get the iterative scheme in motion.

Table 14. Problem difficulties and regression lines

Item	Anagrams	Mill Hill	Numbers	Raven
1	−0.64	−0.59	−0.66	−0.49
2	−0.60	−0.23	−0.07	0.28
3	0.55	−0.25	−0.29	0.55
4	1.50	−0.38	0.77	−0.43
5	−0.40	−0.12	−0.62	0.22
6	0.45	0.33	1.73	0.67
7	0.07	1.34	0.88	−0.32
8	0.97	0.27	1.30	0.14
9	1.19	0.61	0.53	1.53
10	0.76	0.63	−0.18	−0.14
11	0.68	1.14	0.74	0.00
12	0.26	1.84	0.84	0.80
13	1.76	1.49	1.46	–

Problem set	Intercept	Slope	Correlation
Anagrams	−1.01	3.11	.99
Mill Hill	− .83	2.79	.97
Numbers	− .96	2.95	.99
Raven	− .55	3.34	.95

We find the similarity in regression lines rather striking. The similarity is even more striking when we inspect the corresponding correlation coefficients displayed to the right of the regression coefficients. It seems to suggest that there may well be closed-form expressions for the optimal difficulty values. We have not, however, been able to derive such expressions.

Since we have not derived any distribution theory for our loss function we have no test of significance for goodness of fit of the model. At the moment the loss function is but a tool to enable us to fit the model. Since the separate optimization problems involve differing numbers of problems and, as well, differing numbers of subjects, these values (for the loss function) are not even directly comparable from one analysis to another.

Since no alternative model exists which utilizes the same observations we have no basis for a direct comparison with other models. Instead, we proceed to investigate

Table 15. Correlations with basic statistics

	Proportions			Mean Time		
Speed	A	R	W	A	R	W
Anagrams	.02	.17	−.17	−.08	−.24	−.09
Mill Hill	.07	.27	−.31	−.07	−.34	−.08
Numbers	.23	.25	−.39	.13	−.17	.20
Raven	−.15	.02	.03	−.18	−.50	−.30
Accuracy						
Anagrams	.02	.85	−.79	.31	.20	.30
Mill Hill	.00	.94	−.91	.12	−.40	.26
Numbers	.24	.85	−.90	.31	.14	.51
Raven	.01	.96	−.97	.07	−.13	.22
Persistence						
Anagrams	−.85	.52	.33	.70	.45	.37
Mill Hill	−.97	.38	.26	.27	−.06	−.21
Numbers	−.89	.43	.37	.65	.18	−.21
Raven	−.71	.37	−.13	.78	.04	.58

A ≡ abandon; R ≡ right (correct); W ≡ wrong (errors + abandonments)

individual differences in the estimates for speed, accuracy, and persistence which have been derived for the subjects.

2. In Table 15 we display some correlations between our subject's speed, accuracy, and persistence scores and eight simple statistics derived, where possible, for each subject. Each correlation is based on those subjects for whom both measures involved in the particular correlation exist. Thus, the sample sizes for the different correlation coefficients differ from correlation to correlation. Since none of the analyses which follow are based on these correlations we feel justified in neglecting to report the sample size for individual coefficients.

We comment briefly on some salient features of this matrix, in which the columns correspond, respectively, to the proportion of problems abandoned, proportion of problems correct, proportion of problems incorrect, mean time to abandonment, mean time to correct response, and mean time to incorrect response. While these statements are not very striking they do ring a certain face validity and they do orient us towards the analyses which follow.

Subjects with high persistences scores abandon fewer problems than do subjects with low persistence scores. Subjects with high persistence scores have higher mean time to abandonment than do subjects with low persistence scores. Subjects with high persistence achieve more correct responses than do subjects with low persistence scores. Subjects with high accuracy scores achieve more correct responses than do subjects with low accuracy scores. Subjects with high accuracy scores make fewer 'errors or abandonments' than do subjects with low accuracy scores. Subjects with high accuracy scores have higher mean time to 'error or abandonment' than do subjects with low accuracy scores. Subjects with high speed scores have lower mean time to correct response than do subjects with low speed scores.

3. In the top part of Table 16 we display three correlation matrices. Each is based on data from all subjects for whom scores exist on all of the variables involved. The sample sizes for these analyses have already been given in Table 13.

4. For each subject involved in this matrix

Table 16. Intercorrelation matrices

Speed					Factor loadings
1.00	.15	.44	.17	$\chi^2 = 4.29$	.56
	1.00	.20	.26	d.f. = 2	.30
		1.00	.29	$p =$ .12	.76
			1.00		.39
Accuracy					
1.00	.47	.29	.39	$\chi^2 = 4.11$	.58
	1.00	.53	.41	d.f. = 2	.81
		1.00	.38	$P =$ .13	.64
			1.00		.56
Persistence					
1.00	.22	.39		$\chi^2 = 0.00$	.52
	1.00	.31		d.f. = 0	.41
		1.00		P not defined	.75
Speed × accuracy					
.33	.34	.33	.12		
.10	.33	.11	.06		
.19	.39	.36	.28		
−.07	.25	.03	−.13		
Speed × persistence			Accuracy × persistence		
.00	.21	−.06	.11	.15	−.10
−.12	−.01	−.20	.16	.16	.02
.19	.18	−.10	.11	−.07	−.10
−.08	.05	−.07	.11	−.14	.04

we have speed and scouracy scores on each of the four subtests. In this matrix we display the intercorrelations among the four speed measures.

Inspection of this matrix suggests a single factor loading on all four speed measures. Thus, we fitted an unrestricted maximum likelihood model (Joreskog 1967) with but a single factor. The goodness of fit was quite acequate. The factor loadings thus derived appear in the column to the right of the correlation matrix.

5. A similar analysis was carried out on the correlation matrix among the four accuracy scores for these same subjects. Again, the intercorrelations among the four scores are substantially positive and, not surprisingly, once more a single common factor clearly accounts for the intercorrelations.

6. As we have already seen, we would have had to discard most of our subjects to compute the parallel analysis on all four persistence scores. Consequently, we dropped the persistence score on the Ravens subtest and fitted a one-factor model to the intercorrelations on the three remaining persistence scores. Here, once more, a single common factor, loading strongly on all three persistence scores, accounts quite adequately for the intercorrelations among persistence scores.

7. In the bottom part of Table 16 we display the matrix of cross-correlations between the four speed scores and the three persistence scores computed, as above, on all subjects for whom all seven scores exist.

Inspection of this matrix suggests that there is no strong relationship between these two sets of scores. The corresponding canonical correlation analysis which is summarized in Table 17 confirms this impression. Indeed, there seems to be no systematic relationship at all!

8. This analysis parallels the analysis just discussed. The correlation matrix appears in the bottom part of Table 6; and the canonical correlation analysis summarized in Table 17 confirms the very strong impression that, as with the speed and persistence scores, there are no systematic relationships

Table 17. Canonical correlations I

Speed vs persistence			
Canonical correlation	Chi-square	d.f.	P
.34	15.42	12	.22
Accuracy vs persistence			
Canonical correlation	Chi-square	d.f.	P
.36	14.76	12	.26
Speed vs accuracy			
Canonical correlation	Chi-square	d.f.	P
.52	71.76	16	.000
.42	33.3	9	.000
.27	9.93	4	.042
.09	1.05	1	.305

between the accuracy scores and the persistence scores for the subjects.

9. When we inspect the cross-correlations between the speed and accuracy scores, computed once more, on all subjects for whom all eight scores involved exist, a somewhat different impression emerges. Here (with the exception of correlations involving speed score on the Ravens subtest, which seems to be an odd man out) the cross-correlations are predominately positive. The canonical correlation analysis, summarized in Table 17, confirms a very strong relationship between the two sets of measures. It confirms as well, perhaps unfortunately, that this relationship is not unidimensional. Indeed, and probably because of the inclusion of the aberrant speed score on the Ravens subtest, it suggests that at least two dimensions are required to account adequately for the pattern of intercorrelations.

Since we could not discern a psychologically interpretable pattern for a second dimension we decided, quite arbitrarily, to ignore the uninterpretable second dimension rather than to arbitrarily discard the speed score of the Ravens subtest.

10. The analyses discussed above suggest that much, if not most, of the structure in the intercorrelations among the speed and accuracy scores may be accounted for in terms of a factor model involving two correlated non-overlapping group factors loading, respectively, on the speed scores and on the accuracy scores. Consequently, we carried out a restricted maximum likelihood factor analysis using the procedure described by Joreskog (1969).

This analysis yeilded the solution displayed in Table 18 where A denotes the matrix of primary factor loadings and Φ denotes the matrix of interfactor correlations. The third factor was added to account for the intercorrelations among the persistence scores and for the lack of cross-correlation between the persistence scores and both the speed scores and accuracy scores. We note that all non-zero factor loadings are significant in terms of standard error as given by the Joreskog procedure employed but that no overall goodness of fit statistic is available since we have pieced the analysis together from separate analyses. Such a statistic, of course, would be quite superfluous since we know that the submatrix involving the cross-correlations between the speed scores and the accuracy scores cannot be adequately accounted for by such a simple model. We do know as well, however, that each of the remaining five submatrices is adequately accounted for by the simple submodels implied by the matrices in Table 18.

11. It may be readily verified that the factor solution depicted in Table 20 is equivalent to that depicted in Table 18 in the sense that both solutions yield the same expected values for all intercorrelations. The form of

Table 18. Factor analysis

$$\begin{bmatrix} .60 & 0 & 0 \\ .34 & 0 & 0 \\ .72 & 0 & 0 \\ .35 & 0 & 0 \\ 0 & .55 & 0 \\ 0 & .84 & 0 \\ 0 & .64 & 0 \\ 0 & .52 & 0 \\ 0 & 0 & .51 \\ 0 & 0 & .41 \\ 0 & 0 & .75 \end{bmatrix} = A$$

$$\begin{bmatrix} 1 & .68 & 0 \\ .68 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \Phi$$

Table 19. Factor transformation I

$$\begin{bmatrix} .50 & .34 & 0. & 0. \\ .28 & .19 & 0. & 0. \\ .59 & .41 & 0. & 0. \\ .29 & .20 & 0. & 0. \\ .45 & 0. & .31 & 0. \\ .69 & 0. & .47 & 0. \\ .53 & 0. & .36 & 0. \\ .43 & 0. & .29 & 0. \\ 0. & 0. & 0. & .52 \\ 0. & 0. & 0. & .41 \\ 0. & 0. & 0. & .75 \end{bmatrix} = A$$

$$\begin{bmatrix} 1.0 & 0. & 0. & 0. \\ 0. & 1.0 & 0. & 0. \\ 0. & 0. & 1.0 & 0. \\ 0. & 0. & 0. & 1.0 \end{bmatrix} = \Phi$$

Table 20. Factor transformation II

$$\begin{bmatrix} .39 & .39 & .24 \\ .22 & .22 & .14 \\ .47 & .47 & .29 \\ .23 & .23 & .14 \\ .36 & .36 & -.22 \\ .54 & .54 & -.34 \\ .41 & .41 & -.25 \\ .34 & .34 & -.21 \\ .37 & -.37 & 0 \\ .29 & -.29 & 0 \\ .53 & -.53 & 0 \end{bmatrix} = A$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \Phi$$

this three-factor orthorgonal model is somewhat different. It suggests a general factor common to speed, accuracy, and persistence scores, a second factor, orthogonal to the first, which contrasts speed and accuracy scores with persistence scores, and a third factor, orthogonal to the others, which contrasts speed scores with accuracy scores and is not involved at all in persistence scores. This hierarchical model has a sort of superficial validity in common sense terms but there is no one-to-one correspondence with the (speed, accuracy, and persistence) constructs in our conceptual model. To this extent, at least, it seems to be somewhat inadequate.

12. The factor solution displayed in Table 19 was also derived from that displayed in Table 18. Again, it is easy to verify, by direct, multiplication, that the solutions are equivalant in the sense that both yield the same matrix of expected intercorrelations among the scores involved.

Here, however, we have a somewhat different picture. There are now four factors. Factors 2–4 are orthogonal, non-overlapping group factors loading, respectively, on speed scores, accuracy scores, and persistence scores. Factor 1 is orthogonal to each of the others. It is involved in both speed scores and in accuracy scores but not in persistence scores.

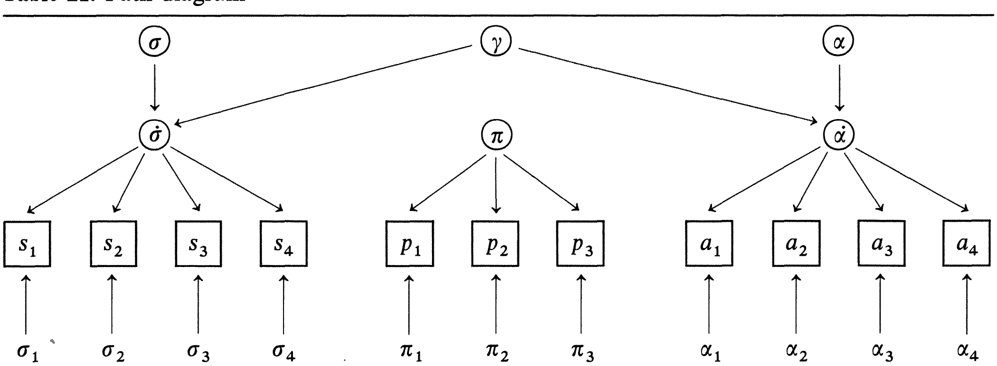
13. In Table 21 we display a path diagram which shows how the factor analyses displayed, respectively, in Table 18 and Table 19 may be related in terms of the analysis of covariance structures (ACOVS) model described by Joreskog (1970). The 'square' boxes represent manifest variables (four speed scores, three persistence scores, and four accuracy scores). The 'circle' boxes represent latent variables.

The latent variables in Table 21 are $\dot{\sigma}$, $\dot{\alpha}$, and π, which denote speed, accuracy, and persistence, and $\dot{\sigma}$ is strongly correlated with $\dot{\alpha}$.

14. The latent variables γ, σ, α, and π correspond to general cognitive ability, speed, accuracy, and persistence. The latent variables in this set, however, are mutually uncorrelated.

We assume that all latent variables and all manifest variables are scaled to zero mean and unit variance and thus that all covariances are product-moment correlations. Thus the path coefficients denoted by directed arrows are standardized regression

Table 21. Path diagram



coefficients and are also product-moment correlations.

The unboxed symbols in the bottom line are specific factors corresponding to the manifest variables. They are mutually orthogonal and each is assumed to be uncorrelated with all latent variables in the model.

Since submatrices in our 11×11 matrix of intercorrelations among the manifest variables have been computed from different subsets of data this matrix is not appropriate for the Joreskog procedure ACOVS.

Consequently, we derived the corresponding path coefficients manually from the solutions presented in Tables 18 and 19. We display the corresponding path diagram in Table 21.

We argue that this analysis provides strong support for an interpretation in terms of three independent, non-overlapping group factors corresponding to speed, accuracy, and persistence and a fourth more general cognitive factor loading speed and accuracy scores but not loading persistence scores.

15. In the top part of Table 22 we look at possible relationships between our latent trait scores – speed, accuracy, and persistence – with the backround variables sex and age.

Table 22. Correlations with sex and age

	Sex	Age
Speed		
Anagrams	−.02	−.02
Mill Hill	−.08	−.08
Numbers	−.10	−.06
Raven	−.19	−.07
Accuracy		
Anagrams	−.05	.06
Mill Hill	.02	−.02
Numbers	.00	−.07
Raven	.04	.02
Persistence		
Anagrams	.03	.05
Mill Hill	.06	.09
Numbers	.08	−.00
Raven	.27	−.14

The correlations shown in Table 22 indicate quite clearly that performance on our tests, at least as summarized by our latent trait scores, is independent of both sex and age.

16. We also acquired scores for each of our subjects on the Eysenck measures psychotism (P), extraversion (E), neuroticism (N), and dissimulation (L).

The relevant correlations are displayed in Table 23. The canonical correlation analyses relating speed scores, accuracy scores, and persistence scores, respectively, to the personality variables are summarized in Table 24.

Table 23. Correlations with personality

	P	E	N	L
Speed				
Anagrams	−.02	.07	.07	−.09
Mill Hill	−.11	−.19	−.05	−.03
Numbers	−.05	.03	.04	−.16
Raven	−.15	.08	.05	−.10
Accuracy				
Anagrams	−.01	−.01	.09	−.08
Mill Hill	.01	−.09	.03	−.20
Numbers	−.00	−.10	.05	−.09
Raven	.01	−.13	−.11	−.21
Persistence				
Anagrams	.02	−.05	−.07	.17
Mill Hill	.02	.04	.16	−.17
Numbers	−.08	.05	−.06	.03
Ravens	.18	.19	−.05	−.08

Table 24. Canonical correlations II

Canonical correlation	Chi-square	d.f.	*P*
Speed vs personality			
.30	18.00	16	.32
Accuracy vs personality			
.39	25.24	16	.07
Persistence vs personality			
.29	11.44	12	.49

17. It seems quite clear from these analyses that performance on cognitive tests, as summarized by the speed, accuracy, and persistence scores derived from our latent trait model, is independent of the personality variables. It is, of course, conceivable that some derived measure such as a speed/accuracy ratio would relate to personality. We have not yet explored such possibly interesting relationships.

This completes our outline of the results of an application of the model to a fairly extensive set of empirical data.

Some Closely Related Models

We now turn our attention, rather briefly, to the work of Van der Ven (1969, 1971) and to that of Wiseman (1975).

In Table 25 we list the most relevant formulae for the models proposed by Van der Ven and by Wiseman. We ignore the several variants of each model on which their proponents have worked and we concentrate on the essential ingredients of the models as they relate to our own work. Though we make them look very simple (the models) we do not, we hope, do them (the models) an injustice.

Table 25. Models by Wiseman and Van der Ven

Van der Ven
$P[\text{reach}]=\sigma$
$P[\text{correct}\mid\text{reach}]=\alpha$
$P[\text{correct, reach}]=\pi$
$\hat{\pi}=\frac{r}{n},\quad \hat{\sigma}=\frac{r+w}{n},\quad \hat{\alpha}=\frac{r}{r+w}$
$ln\frac{\hat{\alpha}}{1-\hat{\alpha}}=ln\,r-ln\,w$
Wiseman
$P[\text{reach}]=\sigma$
$P[\text{omit}\mid\text{reach}]=w$
$P[\text{correct}\mid\text{not omit, reach}]=\alpha$
$ln\frac{\alpha}{1-\alpha}=a-d$
$ln\frac{w}{1-w}=o-r$
$ln\frac{\sigma}{1-\sigma}=s$

Both Van der Ven and Wiseman deal with 'time-limit' tests and both include speed and accuracy as indiuidual difference variables. In addition, Wiseman includes omissiveness as an individual difference or latent trait variable.

For Van der Ven a subject's speed (σ) is defined as the probability that he will reach the problem before the time limit and his accuracy (α), or precision, is defined as the probability that he will sive a correct response siven that he has reached the problem. Assuming a trade-off between the difficulty of the problem and the speed of the subject, which acts in such a way that the (unconditional) probability of a correct response (π) is, for a given subject, constant across problems, Van der Ven derives maximum likelihood estimates for speed and accuracy.

Thus, $\hat{\pi}=\frac{r}{n}$, $\hat{\sigma}=\frac{r+w}{n}$ and $\hat{\alpha}=\frac{r}{r+w}$.

Wiseman, on the other hand, does not assume the 'constancy' hypothesis and he allows for the fact that subjects, despite instructions, do omit or skip items in a time-limit test. He defines three main probabilities: the probability that a subject will reach a problem before the time limit (σ), the probability that a subject will, given that he has reached the problem before the time limit, omit the problem (w), and the probability α that, given that a subject has reached the problem and has not omitted it, will give the correct response. He models these probabilities in terms of the subject's speed (s), accuracy (a), and omissiveness (o) and in terms of the difficulty of the problem (d) and its resistance (r) to omissiveness.

Our own approach and that of Van der Ven developed independently and virtually simultaneously. Given that Van der Ven worked in a true-score framework (Lord and Novick 1968), rather than in the tradition of latent trait theory and that he seems quite clearly not to have followed the Eysenck-Furneaux tradition, the relative convergence of our independent work seems rather striking. The work of Wiseman, while leaning more heavily towards our own approach, draws heavily on both traditions.

About the same time that we were formulating our basic model and Van der Ven was formulating his, Iseler (1970) proposed a model involving three, competing stochastically independent processes, PS (solution), PE (error), and PA (abandonment) leading to respective outcomes S (solution), E (error), and A (abandonment). The processes are competitive in the sense that, as soon as one of them ends, observable performance is terminated by the very outcome associated with that process. At any time, and for each possible outcome, the intensity with which the associated process tends to its termination, given that the process is still ongoing, is identical to the intensity with which the observable solution tends to being terminated by the outcome.

Iseler discussed some assumptions regarding the probability distribution function of time to correct response, he discussed some ways in which this distribution function could be parameterized in terms of subject and problem parameters, and he derived a number of logistic and normal ogive latent trait models for some special cases. He also proposed maximum likelihood procedures for parameter estimation and for assessing goodness of fit of the models.

He seems not to have combined these submodels into a unified model involving speed, accuracy, persistence, and problem difficulty. Nor, it seems, has he followed up his thorough analytic investigation with a programme of empirical work.

During a lengthy correspondence Iseler suggested that our use of the term 'speed' as some sort of regression of problem difficulty on solution time was somewhat at variance with the ordinary use of the term and that our continuing use of the term in this differing sense could lead to confusion. He argued as well that persistence in our model is measured relative to speed rather than in an absolute manner.

The latter criticism seems to imply that speed and persistence are not independent in our formulation. However, in the data to which we have fitted our model this seems not to have occurred. Indeed, as we have just seen, on all four of our tests, speed and persistence are quite independent. In more recent data involving some 1600 subjects and two long tests, the speed-persistence correlations are $-.05$ and $+.03$ (within tests) and $-.08$ and $+.02$ (across tests), and once more there is no apparent dependence between speed and persistence. In our final section we now formulate a stochastic process model which allows for individual differences in speed, accuracy, persistence, and propensity to guess, and which allows, as well, for differences among problems in terms of difficulty level.

A Stochastic Process Model for Individual Differences in Speed, Accuracy, Persistence, and Propensity to Guess

By the time the work reported above had been completed our plans for the next phase of the project were quite clear. Our first aim was to extend the model to allow for possible individual differences in propensity to guess and our second was to come to grips with the competing process formulation proposed so long ago, in this context, by Iseler (1970) and to formulate our extended model along similar lines.

This phase has now been completed. We have not yet devised a computer program to fit the new model to sets of empirical data. Our task in what follows is to present in new model in some detail and, by way of illustration, to apply the model to two sets of simulated data which we generated for the purpose. In closing we discuss a problem concerning consistent estimators for the problem difficulty parameters.

Table 26. Basic ingredients for a stochastic process model

$$\Delta, \kappa, \tau \longrightarrow \alpha, \sigma, \gamma, \pi \longrightarrow l, x, y, z, t$$

$$\left[\begin{array}{l} \Delta\text{, problem difficulty} \\ \kappa\text{, number of response categories} \\ \tau\text{, time limit} \end{array}\right.$$

$$\left[\begin{array}{l} L=l=\begin{cases}1, & \text{time limit}\\ 0, & \text{otherwise}\end{cases} \\ X=x=\begin{cases}1, & \text{correct}\\ 0, & \text{otherwise}\end{cases} \\ Y=y=\begin{cases}1, & \text{abandon}\\ 0, & \text{otherwise}\end{cases} \\ Z=z\text{, confidence rating } 1,2,3,\ldots \\ T=t\text{, completion time} \end{array}\right.$$

$$\left[\begin{array}{ll} \alpha\text{, accuracy} & \theta=\dfrac{\alpha}{\alpha+\Delta} \\ \sigma\text{, speed} & \\ \gamma\text{, propensity to guess} & \rho=\dfrac{1}{\kappa} \\ \pi\text{, persistence} & \beta=\dfrac{1}{\pi} \end{array}\right.$$

In Table 26 we list the basic ingredients in the formulation. The flow chart at the top of the table depicts a set of input-output relationships mediated by a problem solving process.

Input to the system is a particular problem characterized by a problem difficulty level, Δ, the number of mutually exclusive response alternatives or categories, κ, and a time limit, τ, for the problem. The time limit differs from those traditionally employed in that, when applied, it is applied on a problem-by-problem basis rather than on a testwise basis. The problem solving process is characterized by four unobservable latent trait variables which we call accuracy, α, speed, σ, persistence, π, and propensity to guess, γ. Output from the system is the set of random variables l, x, y, z, and t (realizations, respectively, of the random variables L, X, Y, Z, and T).

The discrete binary random variable L takes the value 1 if the time limit has occurred and takes the value 0 otherwise. The discrete binary random variables X and Y are the same as those employed in our previous formulations. The random variable X takes the value 1 if a correct response occurs and takes the value 0 otherwise. The random variable Y takes the value 1 if an abandonment occurs and takes the value 0 otherwise. The random variable T is, as before, completion time and is strictly positive. If a time limit occurs then X and Y are both 0. The discrete random variable Z takes only positive integer values 1, 2, 3,... and is defined only if there is no time limit and no abandonment (i.e. $L=0$, $Y=0$). It is called the confidence rating.

It seems worth noting at this point that these ingredients all have a fairly clear psychological significance and that there are no convenience parameters like arbitrary slope constants or arbitrary scale values for each response alternative for example. Even the auxiliary variables θ, β, and ρ, which we define to simplify subsequent formulae have, as we shall see, a clear interpretation.

We assume a problem solving process (P) which involves three subprocesses – solution (PS), guessing (PG), and abandonment (PA).

The solution process (PS) proceeds at rate σ and leads, on completion, to correct solution (RS) with probability θ and to incorrect solution (WS) with probability $(1-\theta)$.

The guessing process (PG) also leads to one of two mutually exclusive outcomes. It proceeds at rate γ and leads, on completion, to correct guess (RG) with probability β and to incorrect guess (WG) with probability $(1-\beta)$.

The abandonment process (PA) leads, however, to but a single outcome. It proceeds at rate β and it leads, on completion, to abandonment (A) with probability 1.

We cannot, of course, identify each of the five outcomes (RS, WS, RG, WG, and A) implied by this assumed process. We can only observe whether a correct response (R = RS or RG), an incorrect response (W = WS or WG), or an abandonment (A) has occurred. We record as well the completion time (t_R, t_w, or t_A) associated with the particular outcome observed.

We formulate the assumed process as a

model which involves five competing, stochastically independent, processes (PRS, PWS, PRG, PWG, and PA) corresponding, respectively, to the outcomes (RS, WS, RG, WG, and A) implied by the assumed process. We define these stochastic processes in terms of intensity functions (or hazard functions) $h_{RS}(t)$, $h_{WS}(t)$, $h_{RG}(t)$, $h_{WG}(t)$, and $h_A(t)$, which correspond, respectively, to the processes PRS, PWS, PRG, PWG, and PA. In so doing we define, uniquely, the joint probability (p.d.f.) of outcome, or event, (R, W, or A) and completion time (t_R, t_W, or t_A) for the combined process.

We noted above that neither of the outcomes for the solution process (RS, WS) and neither of the outcomes for the guessing process (RG, WG) may be observed but that outcome (A) for the abandonment process is observable.

We did not note that the hypothetical completion times for the independent processes are not observable. Apart from outcome R (RS or RG), W (WS or WG), and A all that we may observe is the completion time associated with the outcome and this is the minimum, T, of the quintuple (T_{RS}, T_{WS}, T_{RG}, T_{WG}, T_A).

Once we formulate our psychological process as a stochastic process model involving independent, competing processes, the formulae for all relevant likelihood functions follow directly from the theory of competing risks as do the formulae for the probabilities (p.d.f.s) of particular outcomes conditional on completion at time $T=t$.

The intensity (or hazard) function, $h_E(t)$, for the process, PE, which leads to outcome E defines, uniquely, the distribution of notional or hypothetical completion times for the process (PE).

In particular it defines uniquely, for the process, its probability distribution function (probability of completion in the interval t to $t+\Delta T$), its cumulative distribution function (probability of completion by time t), and its survivor function (probability of completion after time t).

These four functions are mathematically equivalent. Each defines the properties of the process uniquely. In the present context it seems more natural to work with the hazard (or intensity) function since, in so doing, we model the process directly, in terms of its local behaviour in time, as a function of problem parameters and of subjects' latent trait variables assumed to be relevant to the process.

If at some instant a process leading, say, to outcome E has not yet terminated, and time t has elapsed since onset of the process, the limiting probability that the process will terminate in the next ΔT is given by

$$h_E(t) = \lim_{\Delta T \to 0+} P[t < T \leqq t + \Delta T \mid T > t]/\Delta T$$

Reflecting its application in such diverse areas as actuarial science, demography, vital statistics, renewal theory, and the reliability of systems, $h_E(t)$ is known variously as the force of decrement, force of mortality, age-specific death rate, age-specific failure rate, intensity function, and hazard function.

Since the hazard function is our basic tool in model formulation and since it is not easy to find relevant details in textbooks on statistics, we include for reference the basic re-

Table 27. Distribution of a continuous variable

(1) $\lim_{\Delta X \to 0+} \dfrac{P[x < X \leqq x + \Delta X \mid X > x]}{\Delta X} = h(x)$

(2) $\lim_{\Delta X \to 0+} \dfrac{P[x < X \leqq x + \Delta X]}{\Delta X} = f(x)$

(3) $P[X \leqq x] = F(x) = \int_0^x f(u)\,du$

(4) $P[X > x] = [1 - F(x)] = S(x)$

(5) $H(x) = \int_0^x h(u)\,d(u)$

(6) $S(x) = \exp[-H(x)]$

(7) $F(x) = 1 - \exp[-H(x)]$

(8) $f(x) = h(x)\exp[-H(x)]$

(9) $h(x) = f(x)/S(x)$
$\quad = f(x)/[1 - F(x)]$
$\quad = -\dfrac{d}{dx}[\log S(x)]$

lationships for a random variable X with realization x. We deal with the discrete case in Table 27 and with the continuous case in Table 28. Since most modern writers use $f(x)$, $F(x)$, $S(x)$, $h(x)$, and $H(x)$ to refer, respectively, to the probability density function, cumulative density function, survivor function, hazard function, and cumulative hazard function, we follow suit.

In Table 29 we display, for reference, the main functions derived in formulations involving independent competing processes. We have processes PE_1, PE_2,... leading to outcomes or events E_1, E_2,... and so on. The process leading to a particular outcome (or event) E, say, is defined by the corresponding hazard function $h_E(t)$. The hazard function for the combined process is $h(t)$ and is the sum of the hazard functions $h_{E_1}(t)+h_{E_2}(t)+\cdots$ for the processes involved and $H(t)$ is the cumulative hazard function for the combined process. The joint probability (p.d.f.) of outcome and event $f(E,t)$ is given by Eq. (1). The p.d.f. of completion time is given by Eq. (2) and the unconditional probability of outcome E is given by Eq. (4). Equations (5) and (6), respectively, give the probability of outcome E, conditional on completion time t, $[f(E|t)]$, and the probability distribution function (p.d.f.) of completion time, conditional on outcome E, $[f(t|E)]$. Since all hazard functions which we will define are proportional, completion time is independent of outcome. The reader who requires more background before proceeding may find help in the simple account given by Cox (1962) or in the less simple accounts in David and Moeschberger (1978) and Kalbfleisch and Prentice (1980).

We now define, in Table 30, the competing process part of our formulation.

The solution process, Eq. (3), proceeds at rate σ and leads, on completion, to solution with probability 1, to correct solution, Eq. (1), with probability θ, and to incorrect solution, Eq. (2), with probability $(1-\theta)$.

Table 28. Distribution of a discrete variable

(1)	$P[X=x]=f(x)$
(2)	$P[X\leqq x]=F(x)$
(3)	$P[X>x]=S(x)$
(4)	$P[X=x+1\|X>x]=h(x)$
(5)	$P[X=x+1,\ X>x]$ $=P[X=x+1\|X>x]\,P[X>x]$ $=P[X>x\|X=x+1]\,P[X=x+1]$ $=P[X=x+1]$
(6)	$S(0)=1$
(7)	$f(0)=F(0)=0$
(8)	$f(x+1)=h(x)\,S(x)$
(9)	$F(x+1)=F(x)+f(x+1)$
(10)	$S(x+1)=S(x)-f(x+1)$
(11)	$h(x)=f(x+1)/[1-F(x)]$

Table 29. Basic formulae for independent competing processes

(1)	$f(E,t)=f(E\|t)f(t)$ $=f(t\|E)f(E)$ $=h_E(t)\,e^{-H(t)}$
(2)	$f(t)=h(t)\,e^{-H(t)}$
(3)	$f(E)=\int_0^\infty f(E,t)\,dt$
(4)	$f(E\|t)=\dfrac{f(E,t)}{f(t)}=\dfrac{h_E(t)}{h(t)}$
(5)	$f(t\|E)=\dfrac{f(E,t)}{f(E)}$

Table 30. Hazard functions for stochastic process model - I

Solution process	
(1)	$h_{RS}(t)=\theta\sigma$
(2)	$h_{WS}(t)=(1-\theta)\,\sigma$
(3)	$h_S(t)=\sigma$
Guessing process	
(4)	$h_{RG}(t)=\rho\gamma$
(5)	$h_{WG}(t)=(1-\rho)\,\gamma$
(6)	$h_G(t)=\gamma$
Abandonment process	
(7)	$h_A(t)=\beta=\dfrac{1}{\pi}$

The guessing process, Eq. (6), proceeds at rate γ and leads, on completion, to guessed response with probability 1, to correct guess, Eq. (4), with probability ρ, and to incorrect guess, Eq. (5), with probability $(1-\rho)$.

The abandonment process, Eq. (7), proceeds at rate β and leads to the single outcome, abandonment, with probability 1.

The processes leading to notional times for the respective outcome classes (correct solution, incorrect solution, correct guess, incorrect guess, and abandonment) are assumed to be stochastically independent.

Notional, or hypothetical, completion times for particular outcome classes obey the following rules:

1. Notional times to solution (correct solution or incorrect solution) decrease with increasing speed.
2. Notional times to guessed outcomes (correct guess or incorrect guess) decrease with increasing propensity to guess.
3. Notional times to abandonment increase with increasing persistence.
4. Notional times to correct solution decrease with increasing accuracy and increase with increasing problem difficulty.
5. Notional times to incorrect solution increase with increasing accuracy and decrease with increasing problem difficulty.
6. Notional times to correct guess increase with number of response alternatives.
7. Notional times to incorrect guess decrease with number of response alternatives.

Thus the hypothetical completion times behave quite reasonably in terms of our common sense notions regarding speed, accuracy, persistence, and propensity to guess. Having made this point we now emphasize, very strongly, that we do not observe completion times for which these rules hold. We observe only min $[T_{RS}, T_{RG}, T_{WS}, T_{WG}, T_A]$, we note which of the outcome classes R(RS or RG), W(WS or WG), or A it comes from, and we record the corresponding output duo (R, t_R), (W, t_W), or (A, t_A). We emphasize further that the distributions of observed completion times for the three outcome classes R, W, and A are identical. Indeed each is exponential with mean $(\sigma+\gamma+\beta)^{-1}$. We note as well that this distribution is independent of (i.e. does not involve) α, Δ, and κ.

In Table 31 we realign the hazard functions for the five subprocesses (PRS, PWS, PRG, PWG, and PA) and we generate the hazard functions for the outcome classes R, W, and A. In Table 32 we bring these results together and generate the hazard function $h(t)$ and the cumulative hazard function, $H(t)$, for the combined process.

In Table 33 we display a number of conditional probabilities which follow from our model.

Table 31. Hazard functions – II

Correct response	
(1)	$h_{RS}(t)=\theta\sigma$
(2)	$h_{RG}(t)=\rho\gamma$
(3)	$h_R(t)=\theta\sigma+\rho\gamma$
Incorrect response	
(4)	$h_{WS}(t)=(1-\theta)\sigma$
(5)	$h_{WG}(t)=(1-\rho)\gamma$
(6)	$h_W(t)=(1-\theta)\sigma+(1-\rho)\gamma$
Abandonment process	
(7)	$h_A(t)=\beta=\dfrac{1}{\pi}$

Table 32. Hazard functions – III

Correct response	
(1)	$h_R(t)=\theta\sigma+\rho\gamma$
Incorrect response	
(2)	$h_W(t)=(1-\theta)\sigma+(1-\rho)\gamma$
Abandonment	
(3)	$h_A(t)=\beta$
Combined process	
(4)	$h(t)=\sigma+\gamma+\beta$
(5)	$H(t)=(\sigma+\gamma+\beta)t$

Table 33. Some conditional probabilities

(1)	$P[\mathrm{R}\mid\bar{\mathrm{A}},\bar{G},T=t]=\dfrac{\alpha}{\alpha+\Delta}=\theta$
(2)	$P[\mathrm{R}\mid\bar{\mathrm{A}},G,T=t]=\dfrac{1}{\kappa}=\rho$
(3)	$P[\mathrm{A}\mid\bar{G},T=t]=\dfrac{\beta}{\sigma+\beta}$
(4)	$P[G\mid\bar{\mathrm{A}},T=t]=\dfrac{\gamma}{\sigma+\gamma}$
(5)	$P[G\mid\bar{\mathrm{A}},\mathrm{R},T=t]=\dfrac{\rho\gamma}{\theta\sigma+\rho\gamma}$
(6)	$P[G\mid\bar{\mathrm{A}},\mathrm{W},T=t]=\dfrac{(1-\rho)\gamma}{(1-\theta)\sigma+(1-\rho)\gamma}$
(7)	$P[\mathrm{W},G\mid\bar{\mathrm{A}},T=t]=\dfrac{(1-\rho)\gamma}{\sigma+\gamma}$

We have seen, in Table 26, that the auxiliary variable β is the reciprocal of persistence and is thus an indicator of low persistence. Equations (1) and (2) show that the auxiliary variables θ and ρ also have direct interpretation. The auxiliary variable θ is the probability of a correct response conditional on no guessing, non-abandonment and on completion at time t. Clearly, as γ and β approach 0 our model approaches the well-known simple Rasch model. The auxiliary variable ρ has direct interpretation as the probability of a correct response conditional both on guessing (and thus, conditional on non-abandonment and on completion at time t.

The probability of an abandonment, conditional on no guessing and on completion at time t is given by Eq. (3). It seems striking that this probability is independent of both (i.e. does not involve) the subject's accuracy and the difficulty level of the problem.

Equation (4) shows that the probability of a guess, conditional on non-abandonment and on completion at time t, is independent of problem difficulty and of the subject's accuracy. This seems to be counter-intuitive. Equations (5) and (6) indicate that there is indeed a relationship, but that it works differently for correct responses and for errors, and that in neither case is the relationship counter-intuitive.

The posterior odds on guessing, following a correct response with completion time $T=t$, increases with the difficulty level of the problem and decreases with the accuracy of the subject. The posterior odds on guessing, following an incorrect response with completion time $T=t$, decreases with the difficulty level of the problem and increases with the accuracy of the subject. Both relationships seem to be in accord with intuition.

Equation (7) gives the probability of a wrong guess, conditional on non-abandonment and on completion at time t. It follows that this probability is $(1-\rho)\left(1+\dfrac{\sigma}{\gamma}\right)$ and thus that this probability, for given κ, and hence for given ρ, is an increasing function of the ratio γ/σ. We will exploit this fact shortly when we introduce confidence ratings into the formulation.

Table 34. Log-likelihood for competing process model

(1) $$\begin{aligned} &\ln L[\alpha,\sigma,\gamma,\pi\mid \mathrm{A},\kappa,x,y,t]\\ &=x\ln[\theta\sigma+\rho\gamma]\\ &\quad+(1-x)(1-y)\ln[(1-\theta)\sigma+(1-\rho)\gamma]\\ &\quad+y\ln\beta-(\sigma+\gamma+\beta)t \end{aligned}$$

In Table 34 we display the logarithm of the likelihood function for a single observation from the independent competing processes model. In practice, of course, α, σ, γ, and π (and thus β) carry a subscript i for individuals, Δ and κ (and hence ρ) carry a subscript j for problem, and x, y, t, and θ carry both subscripts. When a set of $j=1, 2, \ldots, n$ problems is administered to a set of $i=1, 2, \ldots, N$ subjects, the log-likelihood function for the set of nN observations is, assuming local independence across problems, and experimental independence across subjects,

$$\ln L=\sum_j\sum_i \ln L_{\mathrm{ji}},$$

where

$$ln\,L_{ji} = x_{ji}\,ln\,[\theta_{ji}\sigma_i + \rho_j\gamma_i] + (1-x_{ji})(1-y_{ji})\,ln\,[(1-\theta_{ji})\sigma_i + (1-\rho_j)\gamma_i] + y_{ji}\beta_i - (\sigma_i + \gamma_i + \beta_i)\,t_{ji}$$

and

$$\theta_{ji} = \alpha_i/(\alpha_i + \Delta_j).$$

We have not attempted to apply this function in fitting the model to empirical data. Indeed, some numerical examples seem to indicate that due to the fact that we do not know whether a correct response, R, is a solution (RS) or a guess (RG) and that we do not know whether an incorrect response, W, is a solution (WS) or a guess (WG) the solution to the resulting likelihood equations is not unique. We do not pursue this matter further here. Instead we return to the result, cited above, that

$$P[\mathrm{W}, G \,|\, \bar{\mathrm{A}}, T = t] = \frac{(1-\rho)}{1+\dfrac{\sigma}{\gamma}}$$

and use it as the basis for the modelling of our confidence ratings.

Under our model a subject who neither abandons the problem nor guesses is sure that he has selected the correct response (otherwise he should have kept working towards solution) and is thus supremely confident. The only thing that can shake this supreme confidence is the probability, conditional on non-abandonment at time t, of an incorrect guess and this, as we have seen, is given by $(1-\rho)/(1+\sigma/\gamma)$. It seems natural, then, to model the confidence ratings as we have done in Table 35, in which the confidence rating, z, follows the geometric distribution with hazard function $h(z)$ and has expectation $1/h(z)$.

Let us assume that a single subject attempts a set of equivalent problems with difficulty level, Δ, and number of response categories, κ, known and let us ask how we might proceed to estimate, for this subject, the unobservable latent trait variables α, σ, γ, and π. We assume that we have $n = \mathrm{R} + \mathrm{W} + \mathrm{A}$ problems where R, W, and A are the numbers, respectively, of correct responses, errors, and abandonments observed and that Σt is the total completion time, disregarding outcome and that $\bar{z}$ is the mean confidence rating on non-abandoned problems.

It is easy to show that the four statistics in the left column of Table 36 have as expectations the expressions displayed in the right column of Table 36. Since $\theta = \alpha/(\alpha + \Delta)$, and since we assume Δ and κ to be known, the expectations involve as unknowns the latent trait variables α, σ, γ, and π. We equate observed values of the statistics with their expectations and derive the moment estimators displayed in Table 37. At the bottom of Table 37 we present an example computed for a set of 1,000 equivalent problems form data simulated for the purpose. The values in parenthesis are the known true values with which the estimates agree very well.

Table 35. Confidence ratings

(1) $h(z) = \dfrac{(1-\rho)\gamma}{\sigma+\gamma} = \dfrac{1-\rho}{1+\dfrac{\sigma}{\gamma}}$

(2) $ln\,L[\sigma, \gamma \,|\, \rho, z] = (z-1)\,ln\left[\dfrac{\sigma+\rho\gamma}{\sigma+\gamma}\right] + ln\left[\dfrac{(1-\rho)\gamma}{\sigma+\gamma}\right]$

(3) $E(z) = \left[\dfrac{1+\dfrac{\sigma}{\gamma}}{1-\rho}\right]$

Table 36. Some statistics and their expectations

$\dfrac{\mathrm{A}}{\Sigma t}$	$\dfrac{1}{\pi}$
$\dfrac{\mathrm{R}+\mathrm{W}}{\Sigma t}$	$\sigma+\gamma$
$\dfrac{\mathrm{R}}{\Sigma t}$	$\theta\sigma+\rho\gamma$
$(1-\rho)\bar{z}-1$	$\dfrac{\sigma}{\gamma}$

Table 37. Example I

$$\hat{\alpha}=\left[\frac{\mathrm{R}}{\mathrm{R}+\mathrm{W}}-\frac{\rho}{(1-\rho)\bar{z}}\right]\left[\frac{(1-\rho)\bar{z}}{(1-\rho)\bar{z}-1}\right]$$

$$\hat{\sigma}=\left[\frac{\mathrm{R}+\mathrm{W}}{\Sigma t}\right]\left[1-\frac{1}{(1-\rho)\bar{z}}\right]$$

$$\hat{\gamma}=\frac{\mathrm{R}+\mathrm{W}}{(1-\rho)\bar{z}\Sigma t}$$

$$\hat{\pi}=\frac{\Sigma t}{\mathrm{A}}$$

$\mathrm{R}=274$	$\hat{\theta}=0.49\ (0.5)$
$\mathrm{W}=281$	$\hat{\sigma}=0.96\ (1.0)$
$\mathrm{A}=445$	$\hat{\gamma}=0.19\ (0.2)$
$\Sigma t=484.83$	$\hat{\pi}=1.09\ (1.0)$
$\bar{z}=12.32$	$\hat{\alpha}=0.97\Delta$
	$\frac{\hat{\alpha}}{\mathrm{A}}=0.97\ (1.0)$

Table 38. A time switch

(1) $I=\begin{cases}1, & t\geqq\tau\\ 0, & t<\tau\end{cases}$

(2) $h(t)=\lambda I$

(3) $H(t)=\lambda(t-\tau)I$

(4) $f(t)=\lambda I\,e^{-\lambda(t-\tau)I}$

(5) $\ln L=\ln(\lambda I)-\lambda(t-\tau)I$

(6) $\hat{\tau}=t_{\min}$

(7) $\hat{\lambda}=\frac{n}{\Sigma(t-t_{\min})}$

(8) $E(t)=\tau+\frac{1}{\lambda}$

(9) $V(t)=\frac{1}{\lambda^2}$

It is tempting to put forth the conjecture that the four statistics are minimal sufficient for α, σ, γ, and π and that our estimates $\hat{\alpha}$, $\hat{\sigma}$, γ, and $\hat{\pi}$ are maximum likelihood estimators. Though we forego this luxury, we point out that, as γ approaches 0, $\bar{z}$ approaches infinity and the estimates for α, σ, and π approach the maximum likelihood estimates for a reduced model which does not include propensity to guess.

We now return to the notion of a possible problem-by-problem time limit which we introduced as one of the basic ingredients in Table 26. In Table 38 we sketch the derivation for a generalized time switch which generates failure times from an exponential distribution truncated below time $T=\tau$. The variable I is an indicator variable introduced for notational purposes only. The variable λ is the (constant) age-specific failure rate associated with the process. Our main concern is with the expected failure time, $E(t)$, and with its variance, $V(t)$. As λ tends towards infinity, $E(t)$ tends towards τ, $V(t)$ tends towards 0, and the device tends towards a time switch for which, at failure time, $t=\tau$.

Suppose, now, that we couple this device to our problem solving process. If none of the notional times T_{RS}, T_{RG}, T_{WS}, T_{WG}, or T_{A} are less than τ, then the time switch wins and our output vector is $l=0$, $x=0$, $y=0$, $t=\tau$, and $z=z$ is undefined. If the minimum notional time, min $[T_{\mathrm{RS}}, T_{\mathrm{RG}}, T_{\mathrm{WS}}, T_{\mathrm{WG}}, T_{\mathrm{A}}]$ is less than τ then the corresponding process wins and $l=0$. If the winning process is the abandonment process then $x=0$, $y=1$, and z is undefined. Otherwise, $y=0$ and x and z follow the definitions given above. We now list all possibilities:

1. Time limit ($l=1$, $x=y=0$, z undefined, $T=\tau$)
2. Abandonment ($l=0$, $x=0$, $y=1$, z undefined, $T=t$)
3. Correct response ($l=0$, $x=1$, $y=0$, $z=z$, $T=t$)
4. Error ($l=0$, $x=0$, $y=0$, $z=z$, $T=t$)

We now assume, as in our last example, that a set of equivalent problems, each with known difficulty level, Δ, and each with known number of response alternatives, κ, has been administered to a single subject and, as before, we assume local independence across responses to different problems. This time we record number right (R), number wrong (W), number abandoned (A), number of time limits (L), and mean confidence rating for non-time-trapped, non-abandoned outputs ($\bar{z}$). In Table 39 we give expected values for four statistics derived from those just given. We equate each

Table 39. Example II

$\frac{R}{R+W}$	$\frac{\theta\sigma+\rho\gamma}{\sigma+\gamma}$
$\frac{R}{R+W+A}$	$\frac{\beta}{\sigma+\gamma+\beta}$
$\frac{L}{R+W+A+L}$	$e^{-(\sigma+\gamma+\beta)\tau}$
$\bar{z}$	$\frac{1+\frac{\sigma}{\gamma}}{1-\rho}$
$R=123$	$\rho=0.5$
$W=125$	$\tau=0.5$
$A=136$	
$L=116$	
$\bar{z}=3.8520$	

$\hat{\theta}=0.49$, $\hat{\sigma}=0.91$, $\hat{\gamma}=0.98$, $\hat{\beta}=1.04$
and, $\hat{\alpha}=(0.96)\Delta$, $\hat{\pi}=0.96$

Table 40. Transformation of the time scale

(1)	$t \geqq 0$	
(2)	$u=\phi_1(t)=\phi_2^{-1}(t)$	ϕ_1, ϕ_2 one-one
(3)	$t=\phi_2(u)=\phi_1^{-1}(u)$	
(4)	$\phi_1'(t)=d[\phi_1(t)]/dt=du/dt$	
(5)	$\phi_2'(u)=d[\phi_2(u)]/du=dt/du$	
(6)	$f(t)\,dt=g(u)\,du$	
(7)	$F(t)=G(u)$	
(8)	$f(t)=\phi_1'(t)\,g[\phi_1(t)]$	
(9)	$g(u)=\phi_2'(u)\,f[\phi_2(u)]$	
(10)	$\phi_1(t)=G^{-1}[F(t)]$	
(11)	$\phi_2(u)=F^{-1}[G(u)]$	

Table 41. Constant hazard functions vs proportional hazard functions

Model I
$h_R(t)=\theta\sigma+\rho\gamma$
$h_W(t)=(1-\theta)\sigma+(1-\rho)\gamma$
$h_A(t)=\beta$
$h(t)=(\sigma+\gamma+\beta)$
Model II
$h_R(u)=[\theta\sigma+\rho\gamma]\,\phi_2'(u)$
$h_W(u)=[(1-\theta)\sigma+(1-\rho)\gamma]\,\phi_2'(u)$
$h_A(u)=\beta\phi_2'(u)$
$h(u)=(\sigma+\gamma+\beta)\,\phi_2'(u)$
$H(u)=(\sigma+\gamma+\beta)\,\phi_2(u)$
$u=\phi_1(t)$, ϕ_1 strong monotonie
$t=\phi_2(u)=\phi_1^{-1}(u)$, $\phi_2'(u)=d\phi_2(u)/du=dt/du$

observed statistic to its expected value and we solve the resultant equations for the values of interest. We simulated a set of data for 500 problems with $\alpha=\gamma=\pi=\Delta=1$ and $\rho=\tau=0.5$ and we fitted the model just outlined to these data. The results are summarized at the bottom of Table 39. Once more, the fitted values agree quite well with the 'true' values listed above. This time, however, we have computed estimates for speed, σ, persistence, π, and propensity to guess, γ, *without recording completion time*. This striking result suggests at least the possibility of large scale testing programmes in which completion times need not be recorded and in which all subjects, despite differences in processing speed, proceed at the same rate of working.

At first blush our simple constant hazard function model (model I in Table 41) seems much too restrictive.

It implies that the distribution of observed completion times is independent of outcome and it implies, as well, that this distribution is exponential with parameter $(\sigma+\gamma+\beta)$.

Suppose, however, that we have a set of completion times, t, which have been generated under model I and that these completion times have been transformed to new values $u=\phi_1(t)$, ϕ_1 one-one.

It is quite straightforward to determine the probability distribution function; $g(u)$, and thus the cumulative distribution function $G(u)$, for the transformed times u. In Table 40 we display, for reference, the basic equations. If we assume that ϕ_1 is strictly monotonic, as well as one-one, it is quite straightforward to show that the transformed times will conform to model II in Table 41, where ϕ_2 is the inverse function which carries u to t (thus, $t=\phi_2(u)=\phi_1^{-1}(u)$). In model II the hazard functions are no longer independent of time. However, they are proportional and thus model II implies that the p.d.f., $g(u)$, and hence the c.d.f., $G(u)$, are independent of outcome. In this case, however, completion times are not distributed exponentially. It is quite straightforward to show that the likelihood equations for model II are identical to those for model I with the exception that each t is replaced by $\phi_2(u)$.

In this sense, our simple constant hazard function model is equivalent to the more general proportional hazard function model. In practice we have $f(t)=\lambda\exp(-\lambda t)$, $\lambda=(\sigma+\gamma+\beta)$, and, thus, $F(t)=1-\exp(-\lambda t)$. However, we have observed completion times u rather than observed completion times t. If we knew ϕ_2, and thus ϕ_1, we could compute $t=\phi_2(u)$ and fit the simple model I to these transformed times, t.

The generalized Weibull distribution discussed by David and Moeschberger (1978) has been found useful for the analysis of failure times in a variety of disciplines. Suppose, for example, we assume observed completion times, u, to conform to the generalized Weibull distribution with $c_1=2$ and $c_2=0$. It follows, from the development sketched in Table 42, that, to fit model II with $\phi_2'(u)=2u$ to the observed completion times u, we fit the simple, constant hazard function, model I to the transformed completion times, $t=\phi_2(u)=u^2$.

Table 42. A family of transformations

(1)	$\phi_1(t)=t^{\frac{1}{c}}+c_2=u, \quad (c_1>0,\ c_2\geqq 0)$
(2)	$\phi_2(u)=(u-c_2)^{c_1}=t$
(3)	$\phi_2'(u)=c_1(u-c_2)^{c_1-1}=dt/du$
(4)	$f(t)=\lambda\exp(-\lambda t), \quad \lambda>0$
(5)	$g(u)=\lambda c_1(u-c_2)^{c_1-1}\exp[-\lambda(u-c_2)^{c_1}]$
(6)	$u\sim \mathrm{W}:\lambda^{-1};\quad c_1, c_2$
(7)	$h(u)=\lambda c_1(u-c_2)^{c_1-1}$
(8)	$H(u)=\lambda(u-c_2)^{c_1}$

The maximum likelihood estimates are invariant under strong monotonic transformation of the time scale. A necessary and sufficient condition for model II, with $\phi_2'(u)=2u$ to hold for the observed completion times u, is that model I holds for the transformed completion times $t=u^2$.

Another possible solution is to select a fairly general functional form for $g(u)$, and thus for $G(u)$, and to fit the general distribution to the observed times u. Having thus determined $G(u)$, we compute $\phi_2(u)=F^{-1}[G(u)]$ and fit the simple model I to the transformed completion times $t=\phi_2(u)$. Thus, in principle at least, our simple model is more general than it seems at first to be.

One problem with our formulation is that each time we add a new subject, i, to our sample we add a new set of latent trait variables α_i, σ_i, γ_i, and π_i to be estimated. A consequence of this is that the unconditional maximum likelihood estimates for the problem parameters, Δ_j, may not be consistent. We can show that if we eliminate individual differences in propensity to guess, γ_i, then for the reduced model

$$[h_R(t)=\theta\sigma,\ h_W(t)=(1-\theta)\sigma,\ h_A(t)=\beta]$$

consistent conditional maximum likelihood estimates exist for the problem parameters Δ_j. However, we have not been able to do this for the more general model which allows for individual differences in propensity to guess.

If, however, we assume a functional form for the distribution of latent traits in the population, then we can determine consistent maximum likelihood estimates for the problem parameters Δ_j.

Our current work assumes a multivariate normal distribution for latent trait variables α_i, $\ln\sigma_i$, $\ln\gamma_i$, and $\ln\pi_i$. Thus, accuracy, α_i, has a normal distribution and speed, σ_i, propensity to guess, γ_i, and persistence, π_i, have a log normal distribution. It remains to be seen whether the massive computations involved in this exercise will be feasible, in practice, or whether we will have to settle for some more simple compromise. Time will tell.

We find that judicious choice of mathematical symbols makes our definition equations easier to communicate, and makes it

Table 43. Revised notation

σ, speed	(sigma)
α, accuracy	(alpha)
π, persistence	(pi)
γ, guessing propensity	(gamma)
Δ, difficulty lend	(delta)
κ, number of response categories	(kappa)
τ, time limit	(tau)
$R=r=\begin{cases}1, & \text{correct response}\\ 0, & \text{otherwise}\end{cases}$	(IR)
$W=w=\begin{cases}1, & \text{incorrect response}\\ 0, & \text{otherwise}\end{cases}$	(IW)
$A=a=\begin{cases}1, & \text{abandonment}\\ 0, & \text{otherwise}\end{cases}$	(IA)
$L=l=\begin{cases}1, & \text{time limit}\\ 0, & \text{otherwise}\end{cases}$	(IL)
$C=c=1,2,3,\ldots$, confidence rating	(IC)
$T=t$, completion time	(T)
n, number of problems	(NP)
N, number of subjects	(NS)

Table 44. Log-likelihood for single observation with no time limit

(1) $$\begin{aligned} &ln\, L[\sigma,\alpha,\pi,\gamma,\Delta\,|\,\kappa,r,w,a,c,t]\\ &=r\, ln[\theta\sigma+\rho\gamma]+w\, ln[(1-\theta)\,\sigma+(1-\rho)\,\gamma]\\ &\quad+a\, ln[\beta]+(1-a)(c-1)\, ln\left[\frac{\sigma+\rho\gamma}{\sigma+\gamma}\right]\\ &\quad-(\sigma+\gamma+\beta)\,t+(1-a)\, ln\left[\frac{(1-\rho)\gamma}{\sigma+\gamma}\right]\end{aligned}$$

(2) $$ln\frac{\theta}{1-\theta}=\alpha-\Delta$$

(3) $$\beta=\frac{1}{\pi}$$

(4) $$\rho=\frac{1}{\kappa}$$

(5) $$\sigma,\pi,\gamma>0$$

easier to interpret both derived results and computational formulae. We choose our acronyms for the corresponding FORTRAN variables very carefully as well. Resultant programs are less cluttered with comments, much easier to follow, and thus much easier to maintain. It seems fitting, in closing, to display in Table 43 a revised notation which has evolved during the preparation of this document.

We display, as well (Table 44), in the new notation, the logarithm of the likelihood function for a single observation. Since we have no data collected according to the paradigm implied by our second example we ignore, here, the possibility of a time limit on each problem. We note that the model definition (Tables 30–32) and the formulae for conditional probabilities (Table 33) are consistent with the revised notation and thus require no translation. We thus document the notation which we currently employ in our formulations.

It is very striking that our formulation leads to the following apparently implausible consequences regarding outcome and observed completion time:

1. Completion time is independent of outcome
2. Completion time is independent of problem difficulty level
3. Probability of abandonment is independent of problem difficulty level
4. Probability of guessing is independent of problem difficulty level
5. Probability of a time limit is independent of the difficulty level of the problem
6. Confidence rating is independent of problem difficulty level

Consequence (1) stems from the fact that the hazard functions for the competing processes are proportional while consequence (2) stems from the fact that $h_s(t)$ is independent of problem difficulty, Δ. Consequences (3)–(4) follow directly from the fact that notional times to solution are thus independent of problem difficulty level.

It remains to be seen whether these consequences are consistent with empricial data or whether the formulation requires modification. In the latter case we simply replace σ with σ/Δ. Thus $h_{RS}(t)=\theta\sigma/\Delta$, $h_{WS}(t)=(1-\theta)\,\sigma/\Delta$. Consequently $h_S(t)=\sigma/\Delta$ and notional time to solution has expectation Δ/σ. Speed is the regression of problem difficulty level on notional time to solution. An increase in problem difficulty now leads

Table 45. Some numerical examples

α	1	1.25					
σ	1		1.25				
γ	1			1.25			10^{-9}
π	1				0.75		10^9
Δ	1					1.25	1.25
κ	2					5	5
τ	1						10^9
$E(r)$	0.3167	0.3364	0.3327	0.3327	0.2893	0.1846	0.4378
$E(w)$	0.3167	0.2970	0.3327	0.3327	0.2893	0.4192	0.5622
$E(a)$	0.3167	0.3167	0.2958	0.2958	0.3857	0.3354	0.0
$E(l)$	0.0498	0.0498	0.0388	0.0388	0.0357	0.0608	0.0
$E(t)$	0.2670	0.2670	0.2570	0.2570	0.2536	0.2746	1.25
$E(c)$	4.0	4.0	4.5	3.6	4.0	2.25	10^9
	(1)	(2)	(3)	(4)	(5)	(6)	(7)

to longer completion times, to more time limits, to more abandonments, to more guesses, and thus to lower confidence ratings as well as to fewer correct solutions and more incorrect solutions.

We have prepared a set of numerical examples which illustrate some of the points just made. These appear in Table 45. The first four rows of the table correspond to the latent trait variables (α, σ, γ, and π) while the next three rows correspond to the problem parameters (Δ, κ, and τ). The next four rows correspond to expected values for proportion correct, proportion incorrect, proportion abandoned, and proportion of time limits. The final two rows of the table correspond, respectively, to expected completion time and to expected confidence rating.

We start, column (1) with the arbitrary configuration $\kappa=2$, $\alpha=\sigma=\gamma=\pi=\Delta=\tau=1$. The time limit censors some 5% of the completion times. This lowers expected completion time from .3333 to .2670.

In column (2) we increase α to 1.25. This increases $E(r)$, decreases $E(w)$ and leaves everything else unchanged.

In column (3) we increase σ to 1.25. This decreases expected completion time, number of time limits, number of abandonments and number of guesses. It increases expected confidence rating, number of correct responses, and number of errors. It leaves the ratio $r/(r+w)$ unchanged.

In column (4) we increase γ to 1.25. This decreases expected completion time, expected confidence rating, number of time limits, and number of abandonments. It increases number of correct responses and number of errors but, since $\kappa=2$, it leaves the ratio $r/(r+w)$ unchanged.

In column (5) we decrease π to .75. This increases the number of abandonments. It decreases expected completion time, number of time limits, number of correct responses, and number of errors. It leaves both expected confidence rating and proportion $r/(r+w)$ unchanged.

In column (6) we increase π to 1.25 and κ to 5. This increases expected completion time, number of time limits, and number of abandonments. It decreases number of correct responses, increases number of errors and, because of the increase in probability of an incorrect guess, it lowers expected confidence rating.

Finally, in column (7), we eliminate the time limit, the effects of guessing, and the effects of relatively low persistence ($\tau=10^9$, $\gamma=10^{-9}$, and $\pi=10^9$) and we set $\Delta=1.25$ and $\kappa=5$ as in column (6). This eliminates both time limits and abandonments and increases both number of correct responses and errors. It decreases the ratio $r/(r+w)$ because of the increase in Δ. Since γ approaches zero the confidence ratings soar to 10^9 despite the increase in κ. Most striking, perhaps, is the increase in expected

Table 46. Some closed-form estimators

(1)	$\hat{\theta}=\frac{r}{r+w}\left[\frac{(1-\rho)\bar{c}}{(1-\rho)\bar{c}-1}\right]-\left[\frac{\rho}{(1-\rho)\bar{c}-1}\right]$
(2)	$\hat{\sigma}=\frac{\Delta(r+w)}{\Sigma t}\left[\frac{(1-\rho)\bar{c}-1}{(1-\rho)\bar{c}}\right]$
(3)	$\hat{\gamma}=\frac{r+w}{\Sigma t}\left[\frac{1}{(1-\rho)\bar{c}}\right]$
(4)	$\hat{\pi}=\frac{\Sigma t}{a}$
(5)	$\hat{\alpha}=ln\left[\frac{\hat{\theta}}{(1-\hat{\theta})}\right]+\Delta$
(6)	$(1-\rho)E(c)=1+\frac{\sigma}{\Delta\gamma}$

completion time by almost 370% with no decrease in speed of the solution process.

In Table 46 we display closed form expressions for estimators of speed, accuracy, persistence, and propensity to guess for the revised model. Their counterparts have already appeared in Table 37. Here we have $n=r+w+a$ equivalent problems with mean confidence rating $\bar{c}$ and total completion time Σt. As in Table 36 we assume no time limit.

Acknowledgments. The work reported in this chapter received generous support from the Social Science Research Council, and from the Bethlem Royal and Maudsley Hospital Research Fund.

References

Andrich D (1978) A binomial latent trait model for the study of Lickert style attitude questionnaires. Br J Math Stat Psychol 31:84–98

Berger M (1976) A study of the Eysenck-Furneaux approach to the analysis of performance on intelligence tests. Ph D thesis, University of London

Birnbaum A (1968) Some latent trait models and their use in inferring an examinee's ability. Chapters 17–20 In: Lord FM, Novick MR (eds) Statistical theories of mental test scores. Addison-Wesley, Reading MA, pp 395–479

Brierley H (1967) Intelligence measurements. Nature 214:1056

Brierley H (1969) The intellectural speed and accuracy of psychiatric patients diagnosed as neurotic. PhD thesis, University of Leeds

Cox DR (1962) Renewal theory. Methuen, London

David FN, Moeschberger ML (1978) The theory of competing risks. Griffin, London

Eysenck HJ (1953) Uses and abuses of psychology. Penguin, Harmondsworth

Eysenck HJ (1967) Intelligence assessment: a theoretical and experimental approach. Br J Educ Psychol 37:81–98

Eysenck HJ (ed) (1973) The measurement of intelligence. M T P Press, Lancaster

Fischer G (1974) Einführung in die Theorie psychologischer Tests. H Huber, Bern

Furneaux WD (1952) Some speed, error, and difficulty relationships within a problem solving situation. Nature 170:3

Furneaux WD (1955) The Nufferno tests. NFER bull 6:32–36 National Foundation for Educational Research, Slough

Furneaux WD (1960) Intellectual abilities and problem solving behaviour. In: Eysenck HJ (ed) Handbook of abnormal psychology. Pitman, London, pp 167–192

Hanna JF (1969) Some information measures for testing stochastic models. J Math Psychol 6:294–310

Iseler A (1970) Leistungsgeschwindigkeit und Leistungsgüte. Beltz, Weinheim

Joreskog KG (1967) Some contributions to maximum likelihood factor analysis. Psychometrika 32:443–482

Joreskog KG (1969) A general approach to confirmatory maximum likelihood factor analysis. Psychometrika 34:183–202

Joreskog KG (1970) A general method for analysis of covariance structures. Biometrika 57:239–251

Kalbfleisch JD, Prentice RL (1980) The statistical analysis of failure time data. Wiley, New York

Lord FM (1968) An analysis of the verbral scholastic aptitude test using Birnbaum's three parameter logistic model. Educ Psychol Meas 28:989–1020

Lord FM, Novick MR (1968) Statistical theories of mental test scores. Addison-Wesley, Reading Ma.

Parzen E (1960) Modern probability theory and its applications. Wiley, New York

Ramsey JO (1975) Solving implicit equations in data analysis. Psychometrika 40:337–360

Rasch G (1960a) Probabilistic models for some intelligence and attainment tests. Paedagogiske Institut, Copenhagen

Rasch G (1960b) On general laws and the meaning of measurement in Psychology. In: Neman J Proceedings of the fourth Berkeley symposium on mathematical statistics. University of California Press, Berkeley London, pp 321–333

Thorndike EL (1925) The improvement of mental measurement. J Educ Res 11:1–11
Thorndike EL, Bergman EO, Cobb MV, Woodyard E (1927); The measurement of intelligence. Columbia University, Teachers College, New York.
Thurstone LL (1937) Ability, motivation and speed. Psychometrika 2:249–254
Van der Ven AHGS (1969) The binomial error model applied to time limit tests. Schippers, Nijmegen
Van der Ven AHGS (1971) Time limit tests: a critical review. Psychologie (Bern) 26:580
White PO (1973a) A mathematical model for individual differences in problem solving. In: Elithorn A, Jones D (eds) Artificial and human thinking. Elsevier, Amsterdam, pp 235–240
White PO (1973b) Individual differences in speed, accuracy and persistence: a mathematical model for problem solving. In: Eysenck HJ (ed); The measurement of intelligence. Press, M T P Press, Lancaster, pp 246–260
White PO (1976) A note on Keats' generalization of the Rasch model. Psychometrika 41:405
Wiseman M (1975) Omissiveness, personality and intelligence: some latent trait models. PhD thesis, University of London

B Reaction and Inspection Time Measures of Intelligence

4 Reaction Time and Psychometric *g*

A.R. Jensen

Nearly 120 years ago, Sir Francis Galton expressed a theoretical preconception or intuition which most people – certainly most present-day psychologists – would regard as highly *counter*-intuitive, namely, the notion that reaction time (RT) is related to intelligence. The common reactions of disbelief to this notion express the view that nothing as simple, trivial, and nonintellectual as RT could possibly reflect anything as subtle, complex, and mysterious as human intelligence, and it is remarked that the most highly intelligent persons often appear to be slow but deep thinkers. In much of popular thought, *speed* of mental action implies superficiality; *slowness,* profundity.

Although Galton and his immediate successors were unsuccessful in demonstrating the supposed relationship of RT to intelligence and it became common knowledge in psychology that Galton's notion was wrong, it now begins to appear – a century later – that Galton was right after all. That is, his hypothesis was right, or at least partially right; but the means for testing it were inadequate in the late nineteenth and early twentieth century, when the study of RT and intelligence was prematurely abandoned, not to be revived until recent years. Whether Galton was right for the right reason or for the wrong reason, theoretically, cannot be decided in any detail from his sketchy writings on this topic (Galton 1883, 1908).

Before defining any one type of RT more precisely in terms of particular experimental procedures, for RT refers to a class of phenomena, it now seems reasonably safe to conclude, from all of the available evidence, that there is some sort of relationship between RT phenomena and general intelligence as it is measured by our standard psychometric tests. Although there are now quite extensive data linking RT and intelligence, I find it virtually impossible at present to draw any firm conclusion about the true magnitude of the relationship as it would be expressed in terms of a coefficient of correlation. The reason for this uncertainty is mainly twofold: (a) little, if any, RT research has been based on large representative samples of the general population, and (b) virtually no account has been taken of the intertrial ability and day-to-day stability of RT measurements and the use of such information for correcting correlations between RT and intelligence test scores for attenuation. However, that there are statistically *significant* correlations between individual differences in general intelligence and a variety of RT measurements can now hardly be doubted. The general phenomenon presaged by Galton is certainly genuine, even if its general magnitude and theoretical meaning are still obscure.

The fact of a significant relationship between RT and psychometric intelligence has at least two immediate implications for theory and research on intelligence.

First of all, it directly contradicts a widespread conception in contemporary psychology that our current standard tests of intelligence measure nothing but a particular class of specific knowledge and acquired cognitive skills or strategies for dealing with certain types of problems generally considered intellectual. Indeed, intelligence itself is conceived of by many psychologists as consisting of *nothing but* a person's acquired knowledge and skills.

According to this conception, individual differences in intelligence are attributable to differences in opportunities afforded by the environment for acquiring the specific items of knowledge and skills that are called for by the standard tests of intelligence. The contrary view is that the specific knowledge and skills called for by IQ tests are merely a vehicle for measuring individual differences in intelligence, and that intelligence can be described neither adequately nor correctly merely in terms of acquired contents and skills. This is evident not only from our present knowledge of the substantial heritability of IQ, and from the finding of significant correlations between IQ and the latency and amplitude of averaged evoked potentials of the brain, but also from the correlation between RT and IQ. Certain types of RT, which are significantly correlated with IQ, are as completely devoid of knowledge content or cognitive skills, in any acceptable meaning of these terms, as one could imagine for any conscious behavioral act.

Secondly, if there is a correlation between individual differences in RT and intelligence, it seems that research on the much simpler information processing phenomenon, RT, would lead more readily to an adequate theoretical account of it than would attempts to theorize directly about the much more complex phenomenon of intelligence. The theoretical constructs developed to deal with the much simpler instances of information processing exemplified in several distinct RT paradigms might then provide a basis for theoretical formulations about the nature of intelligence. What I have in mind, of course, is the development of potentially falsifiable theories, which are sufficiently limited and specified as to generate emprically testable hypotheses. Attempts to develop a theory of intelligence that are based at the level of the traditional instruments used for the measurement of intelligence, viz., various psychometric tests, seem to have reached a theoretical cul-de-sac, ending with the description of factors and apparently unresolvable arguments over the most appropriate factor model. Three-quarters of a century of factor analytic research with psychometric tests has not led to any generally accepted theory of the nature of intelligence. This is not to say that factor analysis is a useless methodology. Quite the contrary. But its real usefulness is not for the purpose of theory construction itself but merely to help identify and delineate the particular categories or dimensions of individual differences that we wish to investigate with a view to theoretical formulation. It would be a wholly unreasonable and hopeless approach to try to develop a theory to explain individual differences in every single psychometric test item in existence. The fact that test items are intercorrelated to varying degrees means they involve certain common features or processes, whatever these may be, and that items can be grouped or classified according to their degrees of intercorrelation. Factor analysis is the accepted tool for this purpose. If our interest is in the most general ability, which accounts for the intercorrelations among virtually all tests of ability however diverse in external appearance, we should be interested in the best obtainable estimate of the most general factor in the abilities domain. Whether or not one wants to identify this general factor as "intelligence" is really a purely semantic issue and is not worth arguing about. It can best be given the neutral label "psychometric *g*", which I will henceforth refer to simply as *g*. But I hasten to note that *g* is probably more highly correlated with what most people, psychologists and laymen alike, mean by "intelligence" than is any other factor derivable in the abilities domain, and certainly more than any other factor or combination of factors that are orthogonal (i.e., uncorrelated) to *g*. Whether *g* is extracted as the first unrotated principal factor in a common factor analysis or as a higher order factor arising from oblique rotation of the primary factors does not seem to be as crucial an issue as some nonempirical factor analysts would seem to argue. I have yet to see an instance where factor scores based on the first principal factor and on a higher order *g*, when derived from a reasonably sized (ten or more) battery of diverse psychometric tests, were not very highly correlated, usually .90 or above. Therefore, in order to get on with the job of research on "intelligence", I advocate accepting *g*, by either method of factor extraction, as our working definition of it. Admittedly, *g* is not a perfectly determinate and invariant construct. Nor does any single test yield a perfect measure of *g*, even excluding the test's measurement error. But for practical purposes in research on the relationship of RT to *g*, with necessarily limited time for testing each subject, we must resort to one or two good *g* reference tests, that is, tests which have been found consistently to have the highest *g* loadings (after correction for attenuation) in a number of factor analyses with a variety of other tests. Raven's Progressive Matrices (Colored, Standard, or Advanced, for the appropriate level of difficulty) has been our first choice as a practical *g* reference test.

Before reviewing the results of recent studies on RT and *g*, it would be useful to view these RT paradigms in the perspective of all their main predecessors in the history of psychology.

A Chronology of Research on Reaction Time

The study of RT in relation to mental ability has a venerable though spotty history, which is outlined in the following brief chronology of landmark events in the history of mental chronometry, emphasizing particularly those aspects most germane to individual differences in intelligence.

1823: The first important recognition of individual differences in reaction time is credited to the Prussian astronomer F.W. Bessel, who coined the term "personal equation" for the consistent variations among different telescopic observers in recording the exact instant that the transit of a star crossed a hairline in the visual field of the telescope. The need to make corrections for the "personal equation" (i.e., individual differences in reaction time) led to the invention (by a German astronomer, Respold, in 1828) of the chronograph, a device for measuring reaction time (RT) in fractions of a second. A markedly improved chronograph was devised in 1850 by the United States Coast Survey. Since then there has been no real problem in measuring RT with adequate precision in terms of one-hundreths or one-thousandths of a second (milliseconds), although the preelectronic devices were mechanically complicated and cumbersome and required frequent calibration.

1850: Hermann von Helmholtz measured the speed of nerve conduction in frogs and (less accurately) in humans. This discovery was especially important for philosophic as well as scientific reasons. The greatest philosophic intellects of the era, including Immanuel Kant, had declared that mental events would forever be excluded from scientific investigation, which depends on exact measurement, because the basis of mental events is the brain and neurones, and they were postulated to act with *infinite* speed, making their functions therefore unmeasurable. This doctrine was rejected by the nineteenth century physiologists, but, prior to Helmholtz's discovery, their conjectures about the speed of the nerve impulse put it at the speed of light or faster. The leading physiologist of the time, Johannes Müller, claimed the speed of neural transmission to be *sixty times faster* than the velocity of light! Helmholtz found that the speed of neural transmission was actually less than one-third of the speed of sound. The philosophic gap between the mental and the physical was reduced.

1862: Sir Francis Galton was the first to suggest that individual differences in general mental ability could be measured by means of reaction time (RT). Galton was also the first explicitly to conceive of intelligence as a *general* mental ability, anticipating Spearman's *g*. Galton (who was Darwin's half-cousin) believed this general ability to be a product of biological evolution, reflecting Darwinian fitness in the struggle for survival. Since quickness of reaction (and keenness of other elemental sensory-motor functions) would seem to have been advantageous to prehistoric man's survival and evolutionary development, Galton thought that measurements of such functions would provide a good index of general mental ability. He invented a host of such measuring devices, including his own RT apparatus (described in his autobiography 1908, p. 248). It measured only *simple* RT to an auditory stimulus, and was rather too simple and crude to yield sufficiently reliable measurements – the subject simply hit a punching bag with his fist as quickly as possible on hearing a signal. Galton tested literally thousands of persons on this and other sensory-motor tests in his laboratory in the South Kensington Museum of Natural Science. But the results were disappointing. Fellows of the Royal Society, for example, did not perform measurably better than the average run of Londoners.

Although Galton quit his research on mental measurement, an American postdoctoral student, James McKeen Cattel, who spent 2 years working with Galton after receiving his Ph.D. (the first American Ph.D. in psychology) in Wundt's laboratory, carried Galton's ideas about mental measurement back to America, where they fully surfaced in Cattell's laboratory in Columbia University in 1901. Cattell dubbed Galton's various sensory-motor tasks "mental tests" – the first appearance of this term in psychology.

1868: F.C. Donders, a Dutch physiologist, discovered that *choice* RT (i.e., different responses to either of two or more stimuli) is longer than *simple* RT (i.e., a predetermined single response to a single expected stimulus). This observation led to Donders' invention of the *substraction method* of mental chronometry. By subtracting the subject's RT to relatively simple stimuli from the RT to more complex stimuli involving discrimination, choice, and decision, one could measure the higher mental processes involved in the more complex situations. The strictly sensory and motor components in simple RT could be subtracted from choice RT, yielding measurements of the speed of "purely mental" events. This discovery, too, helped in advancing psychology from speculative philosophy to natural science. Much of the essential methodology of recent research in mental chronometry (e.g., RJ Sternberg 1977) represents more sophisticated uses of Donders' subtraction procedure.

1873: Sigmund Exner, an Austrian physiologist, coined the term "reaction time" and discovered the importance of "preparatory set" and the preparatory interval (i.e., the interval between a "warning" or "ready" signal and the reaction stimulus). These procedural factors, he found, affect the variability of RT from trial to trial. If the preparatory interval is not controlled by using a "ready" signal, intraindividual variability in RT is increased. Following Exner, a preparatory signal became standard practice in RT studies.

1885: J Merkel, working in Wundt's laboratory in Leipzig, elaborated on Donders' choice RT experiment and discovered that RT increases quite systematically as a function of the increasing number of choice alternatives in the stimulus and response arrangement. This finding clearly anticipates Hick's law (Hick 1952). Merkel's multiple-choice RT data, as I have plotted them in Fig. 1, nicely illustrates Hick's law. Merkel, of course, did not describe his systematic findings in terms of *bits* of information, for the concepts of information theory and the *bit* as a unit of information were not invented until 1949 (Shannon and Weaver). But the psychological importance of Merkel's finding was that it showed that the time for mental activity (as reflected in RT)

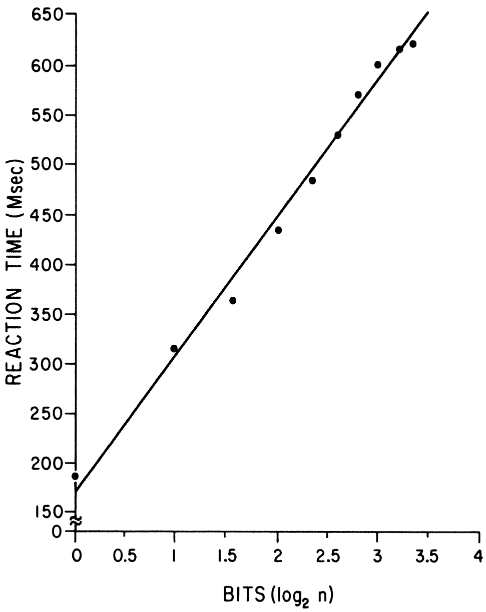


Fig. 1. Mean choice RTs to stimulus arrays conveying various amounts of information scaled in bits; *n* is the number of choice alternatives Data from Merkel (1885) as reported by Woodworth and Schlosberg (1954, p. 33)

is systematically related to the objective complexity of the task.

1894: J Allen Gilbert, at Yale University, was the first to demonstrate what Galton had tried but failed to find – a relationship between RT and "intelligence". Groups of children between ages 6 and 17 who were classified by their teachers as "bright", "average", and "dull" were tested on simple, choice, and discriminative RT. The mean RTs of the three groups were consistently in the same rank order, the "bright" children showing the fastest average RT and the "dull" the slowest. But if IQ test scores had been used and correlation coefficients obtained (neither had yet been invented), probably no statistically significant correlation would have been found. The group mean differences in RT were small (about 20 ms between "bright" and "dull") and the measures were not very reliable. The relationship of RT to intelligence could only have shown up in the differences between the group means, which of course are much less affected by individual measurement error than is a correlation coefficient based on individual scores. Gilbert was also the first to show a very regular, negatively accelerated decrease in RT with increasing age, between 6 and 17 years, and especially a decrease in the trial-totrial intraindividual variability of RT with increasing age. The mean RT of 17-year-olds was almost twice as fast as that of 6-year-olds. This suggests (but does not prove) that RT is related to mental age (a concept that was not to be invented until 1905).

1901: Clark Wissler, working under J McK Cattell at Columbia, was the first to use the coefficient of correlation (invented by Karl Pearson in 1896) to measure the degree of relationship between simple RT and "intelligence" as indexed by the course grades obtained by men students in Columbia College. The correlation was a nonsignificant −.02 – a singularly unimpressive finding. But the deck had been strongly stacked against finding a substantial correlation: each subject's RT was based on an average of only three to five measurements, which we now know would result in exceedingly low reliability; the "range of talent" was highly restricted in this highly selected group of Ivy League students, which we now know greatly attenuates correlations between any *g*-loaded measurements; and the reliability and validity of course grades as a measure of intelligence leave much to be desired. (The best present-day IQ tests show correlations of less than .50 with grades in highly selective colleges.) It was this disappointing result, coming from the then most prestigious psychological laboratory in America, that got into all the psychology textbooks and, for the next three-quarters of a century, cast a pall over the idea of using RT in the study of individual differences in intelligence.

1905: Alfred Binet and Theophile Simon invented the first practically useful intelligence test and conceived of mental age as a means of scaling general ability. This event is important in the history of RT because the Binet-Simon approach to assessing intelligence completely eclipsed the "brass instrument" laboratory techniques for measuring individual differences suggested by Galton and Cattell. For better or worse, no other event has so greatly influenced the whole course of psychometrics as well as present-day theories of intelligence.

1926: H Peak and EG Boring were the first to try to correlate RT with actual intelligence test scores. They insured sufficiently reliable measures of RT by obtaining 100 trials on each subject. Correlations between simple RT and scores on the Army Alpha and Otis intelligence tests were fabulous: −.90 and −1.00, respectively. Unfortunately, these correlations were based on a sample consisting of only five subjects. No one, apparently, was impressed. Peak and Boring (1926, p. 93), however, noted the potential significance of their finding: "...if the relation of intelligence (as the tests have tested it) to reaction time of any sort can finally be established, great consequences, both practical and scientific, would follow".

1927: Vernon Lemmon, working in Cattell's lab at Columbia under Henry Garrett, was the first to find a Pearsonian correlation between both *simple* RT and *choice* RT and scores on an intelligence test (Thorndike Intelligence Test), and he showed that *choice* RT is more highly correlated with IQ than simple RT (−.25 vs −.08) in 100 Columbia College students – a rudimentary demonstration of the relationship of *g* to task complexity. This was the last published study of RT in relation to intelligence until 1964. The low correlation between RT and test scores was an anathema to the psychometric Zeitgeist, which was much more bent on developing tests with practical predictive validity than in experimenting with laboratory techniques for investigating the nature of intelligence.

1949: CE Shannon and W Weaver invented information theory and proposed the *bit* (for *binary digit*) as a measure of information; a *bit* is the amount of information that will reduce uncertainty by one-half. The concepts of information processing theory have played an important role in the subsequent development of mental chronometry as a tool of experimental cognitive psychology.

1952: WG Hick discovered that multiple-choice RT increases as a linear function of the increase in amount of information in the stimulus array, when information is measured in *bits*, that is, the logarithm (to be base 2) of the number of choices. This relationship has become known as Hick's law. The relationship was demonstrated again the following year by Hyman (1953). Hick's law is nicely illustrated by Merkel's (1885) data (as reported by Woodworth and Schlosberg 1954, p. 33), shown in Fig. 1. The slope of this function can be interpreted as a measure of the speed or rate of information processing, expressed as the number of milliseconds per *bit* of information. The reciprocal of the slope (×1000) expresses the rate of information processing in terms of number of *bits* per second.

1964: E Roth, using multiple-choice RTs in an experimental paradigm conforming to Hick's law, found that individual differences in the *slope* of RT as a function of bits (i.e., the rate of information processing) are correlated with IQ. This was probably the first demonstration of a relationship between RT and intelligence that was predicted from the theory that an IQ test measures (among other things) information processing capacity. Individuals differ in the amounts of knowledge and skills called for by ordinary IQ tests, in part, because they differ in the rates with which they process (and hence "acquire") the information offered by the environment. Other things being equal, individuals with greater speed of information processing acquire mor cognitively integrated knowledge and skill per unit of time that they interact with the environment. Seemingly small individual differences in speed of information processing, amounting to only a few milliseconds per *bit* of information, when multiplied by months or yeras of interaction with the environment can account in part for the relatively large differences observed between individuals in vocabulary, general information, and the other developed cognitive skills assessed by IQ tests.

The Hick Paradigm

For convenience, I shall henceforth refer to the general type of procedure used by Roth (1964, see above) as the Hick paradigm, because it is based on Hick's law, that is, the linear increase in RT as a function of the number of *bits* of information conveyed by the reaction stimulus (RS). This paradigm is actually just an elaboration of the simple RT – choice RT (or SRT–CRT) paradigm. The number of choices (n) is merely extended in the Hick paradigm.

Roth's (1964) finding of a relationship between RT (or more exactly the *slope* of RT as a function of bits) and psychometric intelligence, which was first brought to my attention by Eysenck (1967), was the first interesting finding on RT and g in many years. But the encouraging results of this paradigm required replication before we could confidently proceed with it, and that is where I began. Because I now have more information on this paradigm than on any other, I will review in some detail what I and others have learned about it, and point out those aspects which seem the most promising clues for the development of a theory that can account for individual differences (IDs) both in RT and in at least a substantial part of IDs in g – that part of g which can be conceived of as "biological intelligence".

Procedural variations in RT measurement, we have found, have quite important effects on the absolute values of the obtained measurements. It is more doubtful, however, that small procedural variations interact importantly with IDs. Investigators using fairly different RT measurement procedures obtain quite similar relative differences between groups differing in age and intelligence level, and similar correlations with intelligence test scores. But knowledge would cumulate faster in this field if more attention were paid to the procedural aspects and if an attempt were made to make these as uniform as possible, not only for any given RT paradigm, but also across different paradigms. The results of different investigators would be more directly comparable and theoretically useful if there were some more generally agreed upon uniformity of such procedural variables as the modality, intensity, and duration of the preparatory stimulus (PS) and the average length of the (usually) random interval by which the PS precedes the reaction stimulus (RS). The intensity, discriminability, etc., of the RS should also be standardized, when it is not itself the object of experimental investigation. The same strictures should apply equally to the response mode – the type, distance, and strength of movement required for registering the response, the latency of which is the RT. In short, procedural variation should be minimized when it is not the subject of investigation and our chief interest is in IDs. Neither the physical nor biological sciences were able to develop very far without standardized instruments and procedures, and there is no reason to believe that psychology will be an exception. There comes a point in theory development where the absolute values of physical measurements (not just standardized normative scores) that constitute a ratio scale become of crucidal importance, as in direct comparisons (not just correlations) of the periodicity or intraindividual variability of measurements of RT, evoked brain potentials, and critical flicker frequency (CFF).

A Reaction Time-Movement Time Apparatus for the Hick Paradigm. Roth's (1964) RT apparatus and procedures were not very clearly specified. Subjects (Ss) were required to turn off a light as fast as possible after it went on by pressing a button directly adjacent to the light. The amount of information was varied by presenting a different number of light/button alternatives in the array. On each trial only one light in the array goes on. As the particular light that goes on in any given trial is determined at random, the S is kept in complete uncertainty until the instant one of the lights goes on. Only when there is but one light/button in the array is the S confronted with zero uncertainty. Roth's RT measure, that is, the interval between the light's going on and the S's turning it off, includes not only the shortest time it takes for the S to decide to react to the RS, but also the time it takes the S to move his hand through some unspecified distance to push the button which turns out the light. This can be termed movement time (MT) as distinct from RT. RT has also been referred to as "decision time", but the time for any overt act probably includes something more than sheer mental decision time, and so in this behavioral context I prefer the term "reaction time" or RT. But the "RT" in Roth's procedure can be, and should be, experimentally divided into RT and MT. I have devised the RT – MT apparatus to accomplish this and other refinements of Roth's procedure.

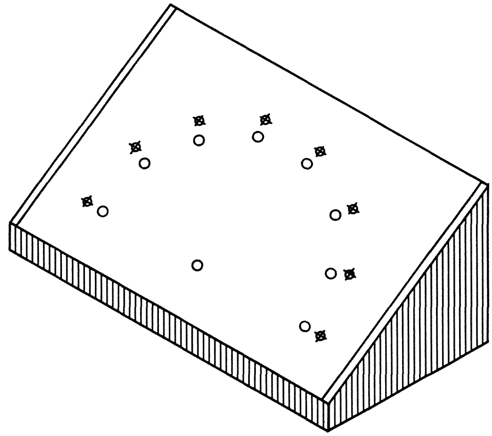


Fig. 2. Subject's console of the RT – MT apparatus. Pushbuttons indicated by circles, green jeweled lights by crossed circles. The "home" button is in the lower center, 6 in. from each response button

The S's console of the apparatus for measuring RT and MT is shown in Fig. 2. It consists of a panel, 13 × 17 in., painted flat black, and tilted at a 30° angle. At the lower center of the panel is a red pushbutton, 1/2 in. in diameter, called the "home" button. Arranged in a semicricle above the "home" button are eight red pushbuttons, all equidistant (6 in.) from the "home" button. Half an inch above each button (except the "home" button) is a $^{1}/_{2}$-in. faceted green light. Different flat black panels can be fastened over the whole array so as to expose arrays having either 1, 2, 4, 6, or 8 light/button combinations.

The subject is instructed to place the index finger of his preferred hand on the "home" button; then an auditory warning signal (the preparatory stimulus or PS) is sounded (a high-pitched tone of 1-s duration), followed (after a continuous random interval [the preparatory interval or PI] of from 1 to 4 s) by one of the green lights going "on", which the subject must turn off as quickly as possible by touching the microswitch button directly below it. RT is the time the subject takes to remove his finger from the "home" button after the green light goes on. MT is the interval between removing the finger from the "home" button and touching the button which turns off the green light. RT and MT are thus experimentally independent. On each trial RT and MT are registered in milliseconds by two electronic timers.

In various studies using the RT – MT apparatus, we have given Ss either 15 or 30 trials, spaced at about 10- to 15-s intervals, on each level of information (i.e., 1, 2, 4, 8 light/button alternatives, corresponding to 0, 1, 2, 3 *bits* of information, where a *bit* is $\log_2$ of the number (n) of alternatives). (Some studies also included six alternatives [or 2.58 bits] in the array.) The levels of information in the array are always presented in their order of magnitude, so the S always begins with the simplest (one light/button) task. Several preliminary practice trials are given to insure that the S understands the task requirements. This has never posed the slightest problem, except in the case of severely retarded Ss, with IQs below 30, who often require more detailed instructions along with demonstration by the experimenter. Under these conditions, Ss with Stanford-Binet IQs as low as 14 have met the task requirements.

Basic Phenomena of the Hick Paradigm. We have now tested about 900 Ss on the RT–MT apparatus, sampled from diverse populations: university students, vocational college students, junior high school and elementary school pupils, borderline mentally retarded in sheltered workshops, and institutionalized mentally retarded. The main expected phenomena of the Hick paradigm have been examined in every set of data. These can be described in general terms for all data sets, noting the few exceptions.

1. *RT and MT as a Function of Bits of Information.* Because the distribution of RT over trials for a single S at any one level of bits is positively skewed, the best measure of the central tendency of RT for an individual is the median RT. This is also true for MT. But the distributions of median RT and median MT over individuals are so nearly normal (although they have a slight positive skew) that we represent the central tendency of groups of Ss by the mean of the individuals' median RTs (or MTs). Woodworth and Schlosberg (1954, p. 37), incidentally, present a graph of the distribution of RT (the average of 30 trials) for 1000 men; it is as perfectly symmetrical and "normal" as one could ever find for any distribution of 1000 physical measurements of any kind.

Figure 3 shows the mean RT and MT as a function of bits for 280 university students. The only statistically significant departure of RT from the linear function known as Hick's law that we have found was in a group of 60 severely retarded adults with a mean IQ of 39. (See group F in

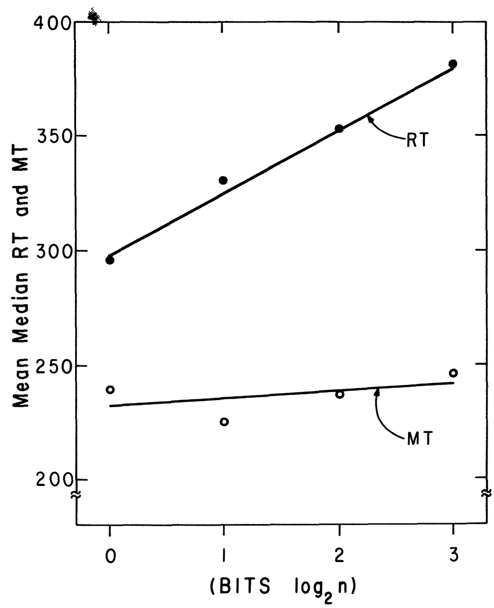


Fig. 3. Mean median RT and MT on the RT–MT apparatus for 280 university students, with 15 trials at each level of bits

Fig. 10.) For borderline retarded and nonretarded Ss, Hick's law is a very robust phenomenon. It is not merely an average statistical effect for a large group of Ss, but appears clearly, with rare exceptions, for individuals when the individual's median RT is plotted as a function of bits. The linear correlation (Pearson *r*) between median RT and bits for individual Ss averages .97, which attests to the close fit of individual RT data to Hick's law.

Hick (1952) suggested calculating bits as $\log_2(n+1)$ instead of $\log_2 n$, where n is the number of alternatives (i.e., light/buttons in the array). Hick reasoned that there are two sources of uncertainty – the uncertainty of which light will go on, which is $\log_2 n$, and the uncertainty as to the precise moment the light will go on. He conjectured that the temporal uncertainty is equivalent to the increase in uncertainty that would result from the addition of one more alternative, i.e., $(n+1)$, and hence bits $=\log_2(n+1)$. However, we have found no consistently better fit to this function in our RT data than to the simpler $\log_2 n$, and so we have used the simpler formulation. The differences in goodness of fit are usually so minute as to be negligible. For example, the data points in Fig. 3 are correlated .996 with $\log_2 n$ and .995 with $\log_2(n+1)$. Conceptually, $n+1$ makes sense, but it seems likely that the uncertainty as to when the RS will occur amounts to something less than the amount of increase in uncertainty that results from the addition of one more alternative to the array of potential reaction stimuli, at least in the present RT – MT procedure, with its short preparatory (random) interval (PI) of 1–4 s. The amount of uncertainty as to when the RS will occur is a function of the PI.

MT in all Ss but the severely retarded is much shorter than RT, a fact which virtually all Ss find very surprising, as it is contradicted by their subjective impressions. This is probably related to the fact that RT is generally faster than the speed of conscious awareness of a peripheral stimulus, which is about 500 ms, as determined by a neurophysiological method involving direct electrical stimulation of the brain (Libet 1965, Libet et al. 1971).

MT always parts company with RT in its relation to bits. MT never shows a significant increase as a function of bits, or in fact any significant or consistent correlation at all with bits. RT and MT clearly seem to involve different processes. Over single trials for an individual S, RT, and MT show zero correlation. That is, there is no correlation whatever between RT and MT (paired over trials) within individual Ss. Individual differences in median RT and median MT, however, are correlated about .40, indicating that they share some common source of variance among individuals. Median RTs for different levels of bits are much more highly intercorrelated than the correlation between median RT and median MT for the same level of bits. The same thing is true for MT. This amounts to saying that IDs in RT and MT involve both a common factor and uncorrelated specific factors, and for this reason it is inadvisable to allow the two variables to be lumped together, as is done in many RT paradigms in both the past and current literature. I suggest that any RT paradigm involving manual response selection should use a "home button" so as to permit the separate measurement of RT and MT.

2. *Reaction Time and Hick's Law Without Response Selection.* I had wondered if Hick's law, as manifested in the RT – MT paradigm, depended on the S's uncertainty of the reaction stimulus (RS) per se or on the task's requirement of response selection. Are the increments in RT with an increasing number of possible response alternatives the result of having to select from among n alternatives the appropriate "program" for the execution of the precise ballistic movement to press the button which turns out the light? We investigated this by having 25 college students do the RT – MT task under two conditions: first, 15 trials under a "single response" condition, then 15 trials under a "double response" requirement. The single response condition only required the S to remove his index finger from the "home button" as fast as possible when the RS (green light) occurred; no other response was called for. The double response condition, which we have routinely used in all other studies, requires the S to remove his finger from the "home button" and press the button adjacent to the light (i.e., the RS) that went on, thus requiring a "double" response – removing the finger from the "home button" and pushing the button 6 in. away, which turns out the light. The results are shown in Fig. 4. Having to make a "double" response adds about 30 ms to the RT and slightly increases the slope of the regression of RT on bits. When the S is required to make the ballistic response to turn out the light, he apparently cannot remove his finger from the "home" button (i.e., RT) until the ballistic response has been "programmed"; the RT under the double response condition thus reflects in part the programming time for the execution of the specific ballistic response required. This outcome is highly suggestive

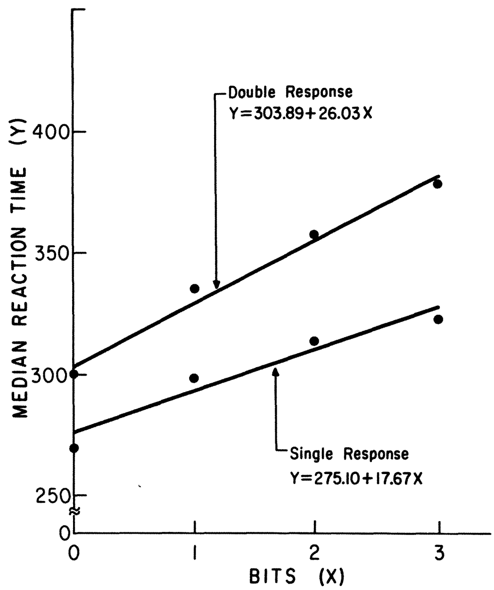


Fig. 4. RT as a function of bits when response selection (i.e., pressing a button to turn off one of the lights) is not required (single response) and when response selection is required (double response)

of Fitts' law, which essentially relates the time for beginning the execution of a movement to the required precision of the movement (Fitts 1954). The ballistic movement programming time of about 30 ms is only slightly affected by the numer of response alternatives. The slope of RT over bits is mainly a function of uncertainty about the RS. But it should not for that reason be thought of as a *sensory* phenomenon per se, for the signal to noise ratio of the RS (a jeweled half-inch diameter green light going on very brightly) is so great as to minimize any between Ss or within S variance due to the discriminability of the RS.

Individual differences in the intercept, slope, and intraindividual variability (over trials) of RT are almost as highly correlated across the "single" and "double" response conditions as the test-retest reliabilities of these variables will permit, and their correlations with psychometric *g* (Raven's matrices) are nearly the same (about −.35) for the two conditions. It seems most likely that *g* is related to the RS uncertainty aspect of RT rather than to the relatively small movement programming component.

3. *Intraindividual Variability in RT and MT.* Surprisingly little attention has been paid to intraindividual variability in the RT literature, with the exception of research on the mentally retarded, which has pointed out that the magnitude of intraindividual variability in RT is one of the most distinguishing features between retarded and normal Ss (Berkson and Baumeister 1967; Baumeister and Kellas 1968a, b, c; Liebert and Baumeister 1973, Wade et al. 1978). In our own work with college students we generally find that intraindividual variability in RT is more highly correlated with *g* measures than is any other single variable that can be derived from the RT−MT paradigm. The reason for the neglect of RT intraindividual variability in most chronometric research is probably that researchers are interested in "goodness" of performance, and the *speed* of RT is a more obvious measure of "goodness" than is the trial-to-trial *variability* of RT. *Speed* of reaction has more the appearance of an "ability" than does *variability* of reaction.

Theoretically, too, variability of RTs would seem to have priority over the average speed of RTs. Assuming an inherent periodicity in the nervous system, the average speed of RT can be seen as a consequence of variability of RT more easily than the reverse relationship.

Intraindividual variability in RT (and MT) is measured by the standard deviation of a S's RTs (or MTs) over trials for any given level of bits, and will henceforth be symbolized as $\sigma_i RT_0$ (or $\sigma_i MT_0$), with the subscript on the RT (or MT) indicating the bits of information conveyed by the RS. The mean of the standard deviations over all levels of bits is symbolized $\bar{\sigma}_i RT$ (or $\bar{\sigma}_i MT$).

Hick (1952, p. 25) claimed that, in his highly practiced Ss, the intraindividual *variance* of RT increases as a *negatively* accelerated function of bits. (This would mean that the *standard deviation* of RT would form

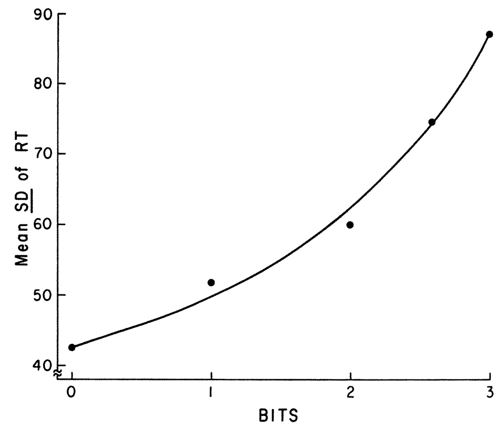


Fig. 5. Mean intraindividual variability (measured by the standard deviation of RTs in milliseconds on 30 trials) as a function of bits on the RT−MT apparatus, for 160 school children in grades four to six

an even more negatively accelerated curve.) In all of our studies, however, intraindividual variability (σ_i) in RT increase as a *positively* accelerated function of bits. Typical results, based on 162 school children in grades 4, 5, and 6, are shown in Fig. 5. A most interesting feature of this curve is that it becomes an almost perfectly linear function if the scale on the ordinate is transformed to a logarithmic scale. Without such a transformation, the function can also be made almost perfectly linear by changing the scale on the abscissa to n (number of alternatives), instead of $\log_2 n$ (=bits). That is to say, RT increases as a linear function of log n, whereas intraindividual variability (σ_i) of RT increases as a linear function of n. This finding will have to be accounted for by any theory of RT.

Intraindividual variability in MT (σ_i MT) is about 1.7 times greater than the average σ_i RT, and, like MT, is completely unrelated to the level of bits. Individual differences in σ_i RT and σ_i MT are correlated only slightly (but significantly) greater than zero, with most rs between about .10 and .20. (These correlations would be raised by about .10 by correction for attenuation.) Also, σ_i MT, very unlike σ_i RT, probably shows the least correlation with g of any of the individual variables derived from the RT−MT paradigm, such as the intercept and slope of RT, $\bar{\sigma}_i$ RT, and even median MT.

4. *The Random Nature of RT Variability.* Intraindividual variability of RT from trial to trial during a single test session displays all the characteristics of random sampling from a population of RTs having a somewhat skewed distribution with a given mean and standard deviation which are characteristic of the S during the particular test session. We know these parameters of RT performance are characteristics of the S, because they show highly reliable IDs within a single test session. However, each S's RTs appear to be generated by a strictly random process, showing a quite consistent variability about the S's mean RT over n trials.

First of all, as would be expected from a random generator, the values of RT show no consistent trend over trials in sessions of 15–30 trials. We have never found a statistically significant practice effect. Dividing trials into first half versus second half yields no greater average difference in RTs or in the σ_i of RT than dividing trials into odd versus even.

Secondly, the covariance matrix of trial-to-trial RTs was tested for homogeneity in a sample of 100 university students. A stringent test of the homogeneity of all of the trial-to-trial covariances in the matrix fails to reject the null hypothesis. (The obtained chi-squared was less than 1/70 th as large as the chi-squared required to reject the null hypothesis at the .05 level of confidence.) In other words, the covariance between any pair of trials does not differ from the covariance between any other pair of trials by more than would be expected from random variation. This is true when there is either 0 or 3 bits of information conveyed by the RS. In other words, the intertrial covariances do not vary more than one should expect if the RTs on each trial represented a sample of one RT drawn *at random* from each of 100 individual distributions having different means and σs. The fact of individual differences is shown by the average intertrial correlation of about +.40. One useful implication of the equivalence of RT from trial-to-trial, except for purely random fluctuation, is that the assumptions of the Spearman-Brown prophesy formula are perfectly satisfied by RT data obtained on a number of trails in a single session.

Although trial-to-trial intraindividual variability of RT meets the two above-described criteria of a random generator, day-to-day variability of the individual median RTs for each daily session, or any other parameter of the Hick paradigm we have examined, such as the intercept, slope, and within-session intraindividual variability, does not meet both criteria of a random generator. Ten Ss tested approximately every other day for nine sessions with 60 trials per session showed no overall average trend in mean RT over the nine sessions (spread over 3 weeks). (An analysis of variance shows nonsignificant F ratios for the main effect of days, i.e., sessions.) But there were slight, statistically significant systematic upward and downward trends for different Ss over the course of nine sessions. The average intercorrelation of RT (median of 15 trials) between days is about +.75, and does not vary as a function of bits. The corresponding MT shows much greater day-to-day stability, with an average correlation of about +.90.

The day-to-day covariance matrix for median RT is not homogeneous, but shows significant variation among the covariances, which form a pattern that approximates a simplex, that is, the largest covariances are between adjacent days or test sessions and they systematically decrease as the number of intervening sessions increases. This simplex pattern of covariances (or correlations) indicates that some form of nonrandom variation in individuals' median RTs occurs over the course of nine test sessions, even though there are no changes in the *average* RT of the group. The same kind of simplex pattern of intercorrelations is usually found for repeated measurements of many other variables that are undergoing gradual and systematic change, such as yearly measurements of children's height and weight, IQ, and trial-to-trial performance on laboratory learning tasks.

Little is known about the sources of day-to-day fluctuations in RT. An individual median RT even fluctuates significantly at different times of the day, and seems to be very sensitive to changes in physiological states associated with eating, sleep cycle, and fatigue. Body temperature fluctuates from hour to hour throughout the day, and RT parallels these temperature fluctuations, higher temperature producing faster RT. Simple RT probably varies about 9 or 10 ms per degree Fahrenheit change in body temperature in the normal range of diurnal variation in temperature. Reviewing this evidence, Woodworth and Schlosberg (1954) note that "the amount of [RT] change [with temperature] corresponds pretty well to what would be expected from the temperature coefficient of chemical processes, and suggests that the cerebral process in reaction depends closely upon chemical activity" (p. 38). It is also of considerable theoretical interest that choice RT shows much larger shifts with change in temperature than does simple RT.

5. *Relationship of RT—MT Parameters to Age of Subjects.* We have examined this in a group of 160 school children ranging in age from 9 to 14 years. Older studies had shown that simple RT has a fairly linear decrease with age between about 5 and 15 years of age, thereafter becoming very negatively decelerated and becoming asymptotic by 17 years of age (e.g. Gilbert 1894). We, too, have found quite linear regressions of RT and MT on age in the range from 9 to 14 years. Thus there is a developmental trend in RT that parallels the developmental trends in physical growth and in other indices of mental development.

Of greater interest to us is the finding that the slope of the regression of mean RT on age increases markedly as a function of the bits of information conveyed by the RS. This is true also for RT σ_i. These results

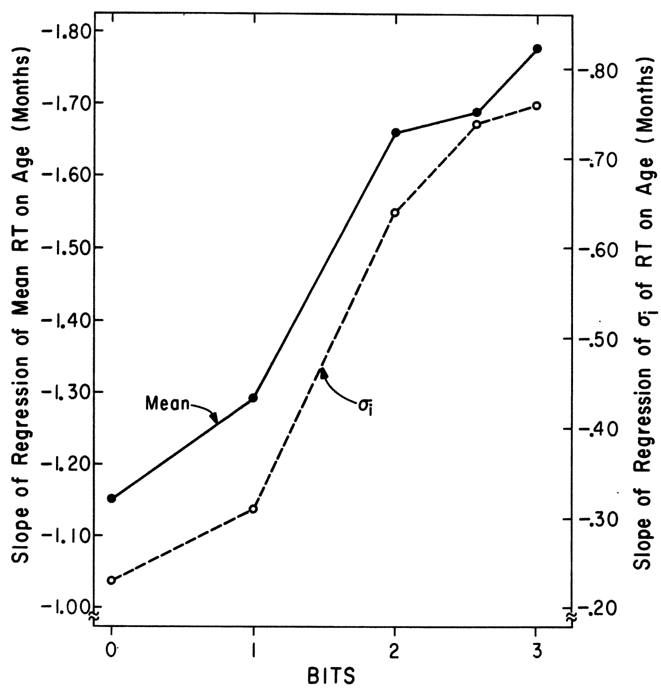


Fig. 6. Slope (i.e., RT decrement in Milliseconds per month of age) of the regression of mean RT and intraindividual variability (σ_i) of RT on age, as a function of bits of information, for 160 children of ages 9–14 years

are shown in Fig. 6. This indicates that performance on the more complex RT tasks (i.e., a greater number of bits) reflects age differences much more sharply than does performance on simpler RT tasks. Children differing 4 years in age, for example, differ about 55 ms in mean RT for simple RT (0 bit), but differ about 85 ms for eight-choice RT (3 bits).

MT shows significant but smaller change with age than RT, the slope of the regression of MT on age being only about 70% of the slope for RT. But more striking is the fact that the regression slopes of mean MT and σ_i MT on age show no relationship to task complexity. Thus, once again MT appears less "cognitive" than RT. Although MT shows a slight but significant developmental trend, it does not seem to be associated with the information processing demands of the task, whereas RT is clearly related to Ss' information processing capacity, which increases much more dramatically than motor speed and accuracy between ages 9 and 14.

RT Paradigms and Psychometric *g*

Consideration of the relationship of RT to psychometric *g* is a complex affair. For one thing, RT is merely a generic term for a great variety of procedures and paradigms for measuring reaction time, and each of these paradigms yields data from which a number of parameters can be derived, such as the intercept, slope, and intraindividual variability, as was pointed out for the Hick paradigm in the previous section. Each of these paradigms and parameters may show correlations with *g*, singly (by Pearson *r*) or in various weighted combinations (multiple *R*).

The correlation coefficient is not necessarily the best or most efficient method for initially discovering which particular paradigm and parameters are related to *g*. Comparison of the *means* of various RT variables obtained in groups that differ in *g* is an efficient exploratory method. Its efficiency, as contrasted with that of correlation analysis, is mainly due to two factors:

1. First is the fact of the day-to-day instability of IDs in RT parameters, especially those most highly related to g. A low stability coefficient, like low reliability in general, puts a low ceiling on the maximum correlation that can be obtained between RT variables and g or any other external criterion measurements. A group's mean, however, is highly stable for all RT parameters. The day-to-day rank order of sample means on RT parameters, provided they are sampled from different populations with respect to the average g of the population, remains highly stable, so that relatively small samples can be used to establish a connection between RT parameters and g. Correlations within any relatively homogeneous group, on the other hand, are highly attenuated by the inherent temporal instability of certain RT parameters and often barely reach significance in samples of less than about 40 Ss. In reviewing the entire literature on various RT correlates of g-loaded tests, the modal Pearson r appears to be somewhere near .35. This much can be said for the correlations, however: virtually never in my examination of this literature, nor in any of our own work, have I come across any $RT \times g$ correlations, whether statistically significant or not, that were on the "wrong" side of zero. That is, the correlations, although often unimpressive, are always in the theoretically expected direction, namely, higher g predicting faster overall RT, lower intercept, and less slope of RT when complexity of the RS is varied over two or more levels, and smaller intraindividual variability in RT over trials. If there have been surprises in this field, they have been due to finding significant and replicable correlations where they were not expected in terms of our earlier theoretical conceptions – for example, the quite pronounced relationship of MT to g in normal children and retarded adults.

2. Second is the fact that investigators are rarely in a position to obtain random or representative samples of the general population. Almost every study I have found in the literature on RT and intelligence, including all of my own studies, have used samples drawn from quite restricted populations with respect to general intelligence. Almost any "natural" group from which one may draw a sample represents some restricted range of the total distribution of IQ in the general population. Restriction of the "range-of-talent", as is well known, plays havoc with correlations. Corrections of the obtained correlations for restriction of range are questionable without highly reliable estimates of the variances of the correlated variables in the general population. Some investigators have made up "artificial" or ad hoc samples composed of individuals selected over a very wide range of IQs, from retarded to gifted. But these "artificial" groups do not represent a sample of any population, and the distribution of IQs within them is usually rectangular (i.e., nearly equal frequencies at every level of IQ), or even bimodal. Correlations between RT and IQ based on such ad hoc samples are usually very high. Their one and only important feature is their statistical significance, for the magnitude of the r is not generalizable to any real population, including the general population, in which the full range of g has an approximately Gaussian frequency distribution. Representing nearly the full range of g found in the general population by a sample with a rectangular distribution, of course, greatly exaggerates the true correlation in the population. Therefore, in our research we prefer to report the raw correlations found within samples of "natural" populations, however restricted in range of IQs, and to observe mean differences in RT parameters between "natural" groups that happen to differ in mean level of IQ.

Because of these complications, it is practically impossible at present to conclude just what the correlations between RT variables and psychometric measures of intelligence might be in the general population, except to say that there is undoubtedly a true correlation between the two classes of variables and the population correlations are probably larger than those found in more restricted "natural" groups. However, the more important point at this stage, from a theoretical standpoint, is that a significant relationship exists between RT phenomena and *g*. That is the primary basis for further investigation.

When I began researching the correlation between RT and *g*, and reviewed the quite sparse literature on this topic, with its significant but usually modest correlations, mostly in the .30 to .40 range, I naturally wondered if there was the risk that these few reports were merely instances of Type I error, and that failures to reject the null hypothesis with respect to RT × *g* correlations had simply not found their way into the published literature. However, I now have very little doubt that our knowledge of the RT × *g* correlation could not be merely Type I error due to the failure of investigators to report negative or insignificant results. In recent years a number of researchers have reported quite consistent results from different RT paradigms. Also, in our own research on the Hick paradigm with a wide variety of groups from different parts of the IQ distribution, we have always found a statistically significant relationship, invariably in the predicted direction, between certain RT (and MT) paramters and mental test scores.

As I have already reviewed the research relating RT paradigms and parameters to psychometric intelligence in some detail elsewhere (Jensen 1980, 1981), I will here only briefly summarize the main findings obtained with different RT paradigms, using graphs to highlight the most telling results.

Simple and Choice RT. Comparison of simple (SRT) and two-choice (CRT) reaction times is probably the simplest of the RT paradigms. CRT is invariably longer than SRT, and usually CRT is the more highly correlated with *g*. Developmental trends from childhood to adolescence are also more pronounced for CRT than for SRT. These findings are typically illustrated in Fig. 7, from a study by Keating and Bobbitt

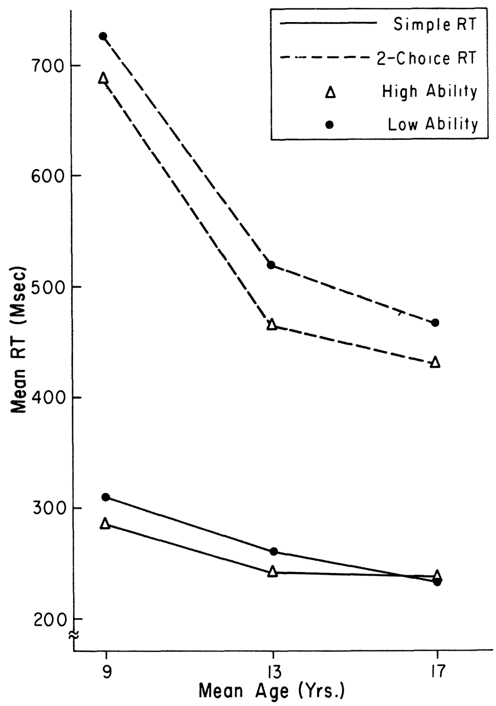


Fig. 7. Simple and choice RT as a function of age and ability level (Raven's Matrices). Keating and Bobbitt (1978)

(1978). SRT required the S to press a button when a red light went on; in CRT the S pressed either a red or a green button when either a red or a green light appeared (in the same aperture). Low and high ability groups were selected from the 40–45 and 90–95 percentiles, respectively, on Raven's Matrices. In this study there was no attempt to distinguish between RT and MT: both variables are amalgamated in the nominal RT, which, therefore, is not directly comparable in absolute magnitudes to the RT obtained in the Hick paradigm using the RT–MT apparatus. The form of the relationships of SRT and CRT to age and ability level, however, is typical.

Hick Paradigm. The typical findings for SRT and CRT extend to the more complex Hick paradigm, which further magnifies the increased relationship of RT to *g* as the complexity of the reaction stimulus is increased. This generalization, which is repeatedly supported by our own research on the Hick paradigm, using the RT – MT apparatus, is most clearly illustrated in a study by Lally and Nettelbeck (1977) reporting the correlation between choice RT and IQ (in a very heterogeneous group ranging from IQ 57 to 130) as a function of bits or $\log_2$ of the number (n) of choice alternatives, as shown in Fig. 8. The same trends are seen in much more homogeneous groups tested in our laboratory, as shown in Fig. 9. This increase in the correlation between RT and *g* as the complexity of the RS is increased is one of the key phenomena that any theory of intelligence must deal with. The theory must also explain why this generalization holds true only in the lowest range of task complexity, extending perhaps from 0 to 4 or 5 bits of information. The upper limit is not clear. But the increasing relationship between RT and IQ seems not to extend beyond the range of tasks to which RT is

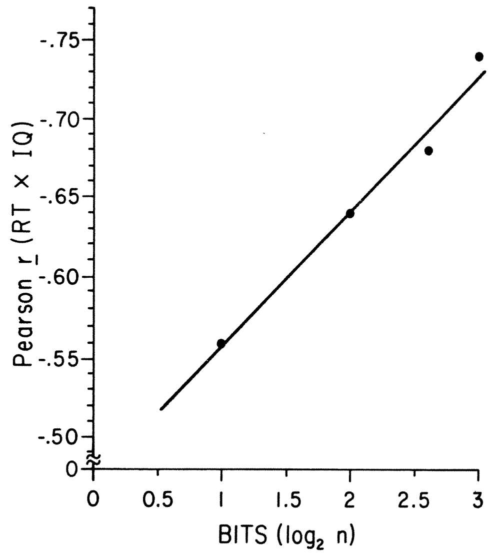


Fig. 8. The correlation (Pearson r) between choice RT and IQ as a function of number of alternatives (n), in a group of 48 Ss with Wechsler Performance IOs ranging from 57 to 130. Lally and Nettelbeck (1977)

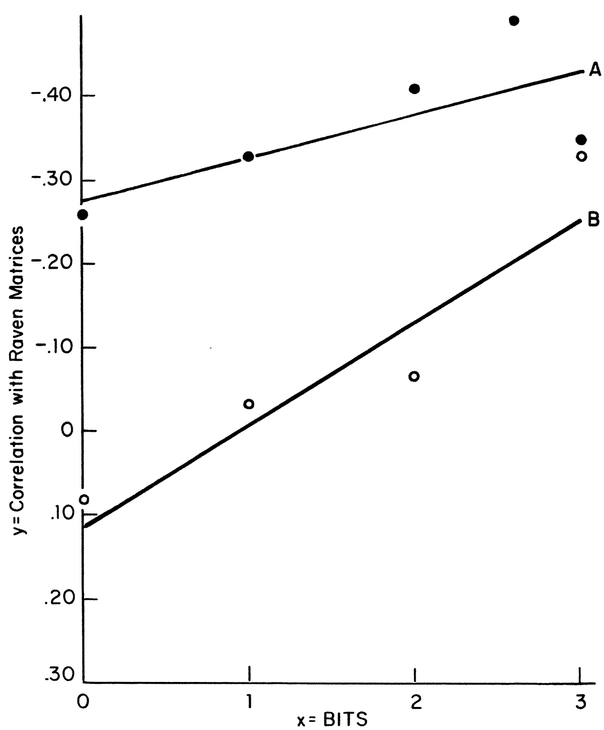


Fig. 9. Correlation (r) of Raven Matrices scores with RT as a function of complexity of the reaction stimulus scaled in bits for (A) 39 female ninth graders (age 14 years) and (B) 50 university students, who, probably because they are more highly selected and consequently more restricted in variability on g, show the smaller correlations

greater than about 1,000 ms. When the processing time is greater than that, further increases in task complexity do not result in a further increase in the RT – IQ correlation (e.g., Spiegel and Bryant 1978). When we measure response time to problems of the degree of complexity of typical intelligence test items that are difficult enough to measure individual differences in terms of number of right and wrong answers under unspeeded condition, the correlation between individual differences in response times and ability as measured by number of items gotten correct on a test usually breaks down completely. For example, the correlation between individual differences in solution times for Raven Matrices items and total score on the Raven has been found to be near zero in three studies (Jensen 1979, Snow et al. 1976, White 1973). I emphasize that the nonsignificant correlations are between (a) individual differences in response times to test items and (b) total scores (i.e., number right) on the test. When solution times for items are *averaged* over Ss, the correlation between mean item solution times and difficulties (i.e., proportion of Ss attempting the item but failing to get the right answer) approaches unity (Elliott and Murray 1977). In other words, more difficult test items (when answered correctly) have longer average response times, but the response times are barely, if at all, correlated with intelligence. I would predict that one would obtain a higher correlation between IQ and response latencies to test items in college students if the test items were from intelligence tests of a difficulty level appropriate for elementary school children than if the items were from ability tests of a difficulty level suitable for college students. I call this the *test-speed paradox*. The explanation of it involves a number of factors.

First, it should be understood that the test-speed paradox holds for test items answered correctly. It would be trivial if it only held for a mixture of right and wrong solutions, as a wrong solution can hardly be expected to reflect all the mental processes that may be necessarily involved in a correct solution. Also, the response times of bright and less bright Ss should be compared on only those items that all Ss get right, otherwise the response times of the brighter Ss would be slower simply because they have solved more difficult items. But beyond these obvious controls, there are other factors that work against a high correlation between test speed and ability, even though, paradoxically, we may find a substantial correlation between test scores and RT parameters derived from relatively simple paradigms in the 0–3 bits range of information processing demands. We know that both intra- and interindividual differences in RTs increase with increasing amounts of information in the RS. However, the *nominal* information in the RS is not linearly related to RT beyond a point. Because of the brain's limited channel capacity, increasing the informational input invokes other processes, such as holding encoded stimuli and partial solutions in short-term memory while performing other operations. So with increasing task complexity, beyond a certain point, the RT departs from linearity. Also it appears that complex tasks requiring considerable time and persistence, such as difficult matrices items, allow personality factors to enter the picture, and these are uncorrelated with ability. We have not found significant correlations between personality variables and performance on relatively simple RT tasks with RTs below 1,000 ms among university students. Yet total time on Raven's Matrices was found to be correlated –.46 with E (extraversion) scores on the Eysenck Personality Inventory, whereas the correlation between total time and Raven scores was exactly zero.

Both the *intercept* and the *slope* of the regression of RT on bits of information in the Hick paradigm are correlated with *g*. This is true when intercepts and slopes are calculated for individuals and when they are calculated for groups of different intelligence levels. In general, the slope parameter seems to be more discriminating for *g* among individuals in more intelligent

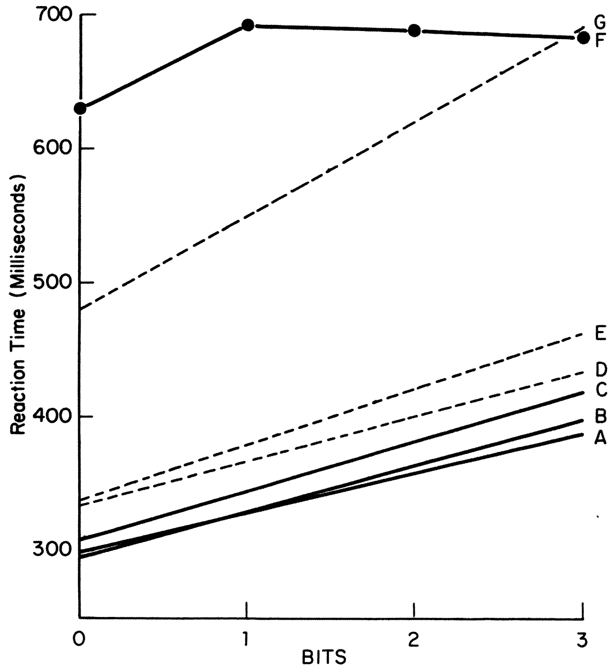


Fig. 10. Reaction time as a function of *bits* in seven different groups: *A*, university students ($N=155$); *B*, ninth grade girls ($N=39$); *C*, sixth graders in a high SES-high IQ school ($N=50$); *D*, *E*, white ($N=119$) and black ($N=99$), respectively, male vocational college freshmen; *F*, severely retarded young adults ($N=60$); *G*, mildly retarded young adults ($N=46$)

groups and the intercept becomes a relatively more important correlate of g in children and retarded adults. One problem with the slope is that it is much less stable from day to day than the intercept. Therefore group comparisons of slope are more informative than correlations between individual measures of slope and g within groups. Figure 10 shows the Hick phenomenon for several groups differing in age and general ability. For all groups except the severely retarded (group F) the data points are omitted for clarity, for in no group except the severely retarded do the data points depart significantly from a linear trend. All of the group in Fig. 10 differ significantly from one another in slope except groups A and B. The two most extreme groups, except for the severely retarded, groups A and G, are shown separately in Fig. 11 and 12. Also shown are the movement time (MT) and the average intraindividual variability (indicated by vertical lines).

When the mildly retarded group in Fig. 12 is split in two at the group's median of the distribution of Raven's scores, we found, to our suprise, that MT discriminates more than RT between the groups, as shown in Fig. 13.

MT also discriminates between IQ levels in a junior high school sample (ninth grade girls), but not as much as RT, as shown in Fig. 14, in which the distribution of Raven scores of the 39 Ss was trichotomized. Note that MT is much faster than RT and MT does not increase significantly over bits.

The only group which is markedly at variance with these general findings is the severely retarded, with IQs ranging from 14 to 60, mean = 39. They fail to manifest Hick's law and it is the one group for which MT is slower than RT, as seen in Fig. 15. In this group, median RT and MT were correlated with g only $-.13$ and $-.18$ respectively, but the $\bar{\sigma}_i$ of RT and $\bar{\sigma}_i$ of MT correlated $-.44$ and $-.57$ (both significant at the .01 level). A simple sum of standardized scores on median RT, median MT, RT$\bar{\sigma}_i$ and MT$\bar{\sigma}_i$, and a measure of "neural adaptability" derived from the average evoked potential were correlated .64 ($P<.001$) with g factor scores based on 15 psychometric tests (Jensen et al. 1981).

The reversal of the speeds of RT and MT

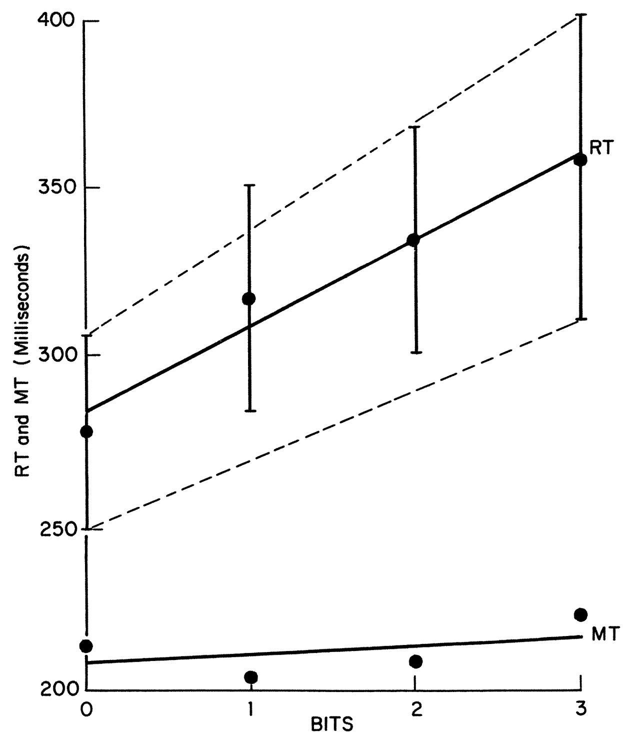


Fig. 11. Mean RT and MT, and the mean $\pm 1\ \sigma_i$ of RT over 15 trials (vertical lines) in 50 university students (group *A* in Fig. 10)

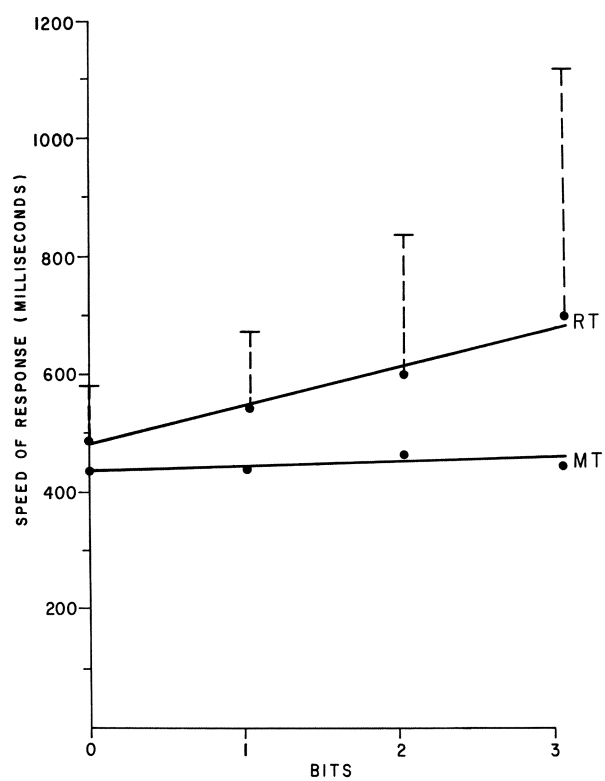


Fig. 12. Mean RT and MT and mean σ_i of RT over 15 trials (*vertical dashed lines*) in 46 borderline retarded young adults (group G in Fig. 10). Vernon (1981)

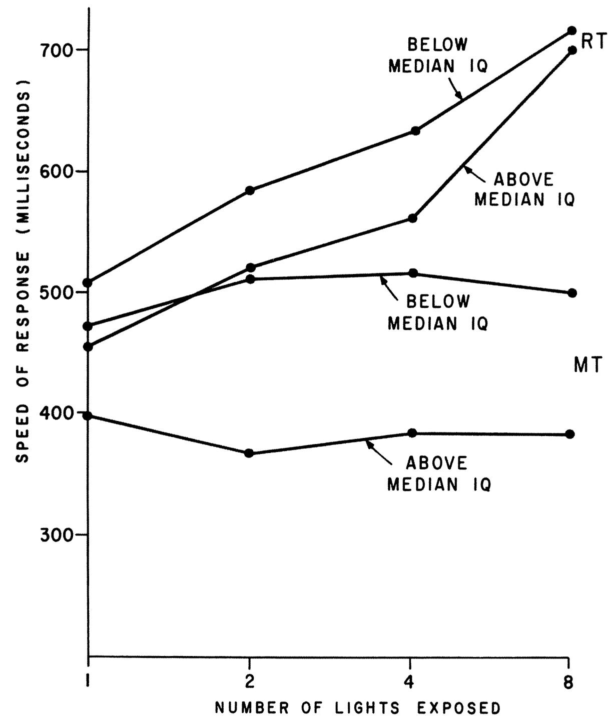


Fig. 13. RT and MT of mildly retarded young adults who are above or below the sample's median IQ (Raven). Vernon (1981)

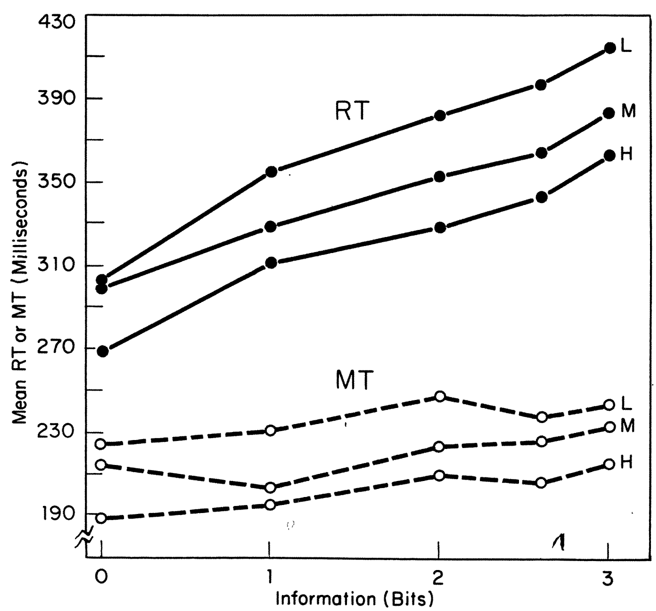


Fig. 14. Mean RT and MT as a function of bits for the high (*H*), middle (*M*), and low (*L*) thirds of the sample ($N = 39$) of ninth grade girls on Raven's Standard Progressive Matrices scores. Jensen and Munro (1979)

in this retarded group caused us to wonder if the ratio of RT/MT bore any relationship to level of intelligence. When the ratio of mean RT/mean MT is plotted for the four adult groups differing in mean IQ, the results show a rather consistent relationship, as seen in Fig. 16. I have hypothesized, in accord with similar findings by Sternberg (1977), that brighter Ss use up relatively more of their RT for "programming" the

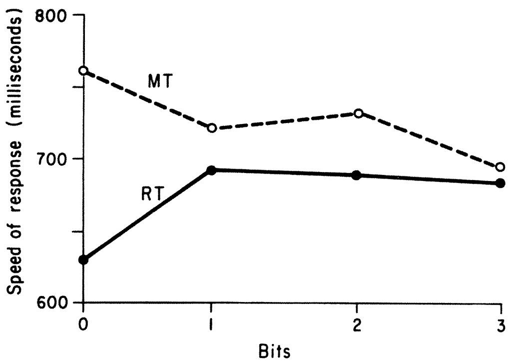


Fig. 15. Mean RT and MT as a function of bits, in 60 severely retarded adults (mean IQ=39). Jensen et al. (1981)

precise ballistic response required to push the button which turns out the light; this lengthens RT relative to MT. Data relevant to this hypothesis are discussed in detail elsewhere (Jensen 1982).

Intraindividual variability ($\bar{\sigma}_i$) in RT, among all of the RT−MT parameters, has generally proved to be the best correlate of *g*. It is the one parameter that shows a significant, and usually the most substantial, correlation with *g* in relatively homogeneous groups at every ability level we have tested from the severely retarded to university students. The *r*s range from about −.30 to −.45 with a mean of −.35, impressive figures considering that RT$\bar{\sigma}_i$ is one of the least stable RT parameters, with a correlation of .42 between RT$\bar{\sigma}_i$ for 100 university students obtained in each of two test sessions 1 day apart. If this represents the typical stability coefficient of RT$\bar{\sigma}_i$, then the average correlation between RT$\bar{\sigma}_i$ and *g*, when corrected for attenuation, would be about −.55. On the assumption that any one group in which the correlation has been determined represents only half of the total variance of *g* in the general population, a correction of the correlation of −.55 for restriction of range on *g* would boost it to about −.70. The true-score population correlation between RT$\bar{\sigma}_i$ and *g* might even be slightly higher than that, because there is undoubtedly also some restriction of range on RT$\bar{\sigma}_i$ in our sample. Mean differences in RT$\bar{\sigma}_i$ between groups, expressed in standard score units (*z*), are almost as large as the mean IQ differences between the groups. For example a university sample and a vocational college sample differ 13 points in IQ and differ 0.68 *z* ($P<.001$) in $\bar{\sigma}_i$ of simple RT.

Individual differences in median RT and

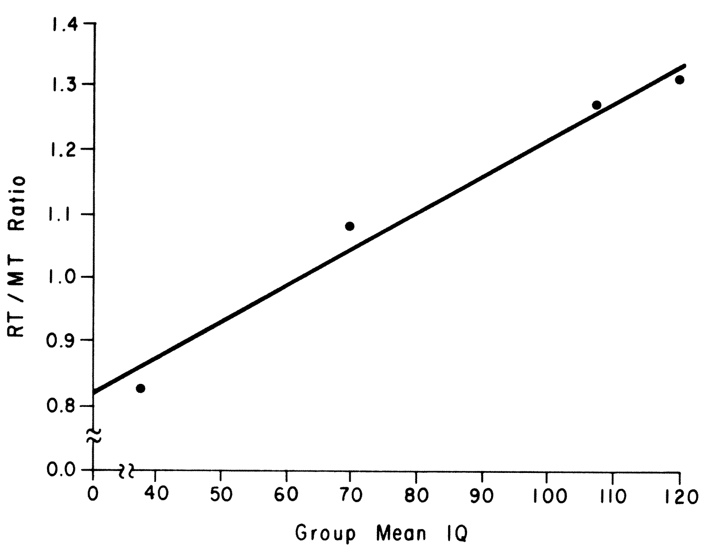


Fig. 16. Mean RT/MT ratio plotted as a function of average IQ levels of four groups: severly retarded ($N=60$), borderline retarded ($N=46$), vocational college students ($N=200$), university students ($N=50$). Mean RT and MT are based only on the one light/button task (0 bits)

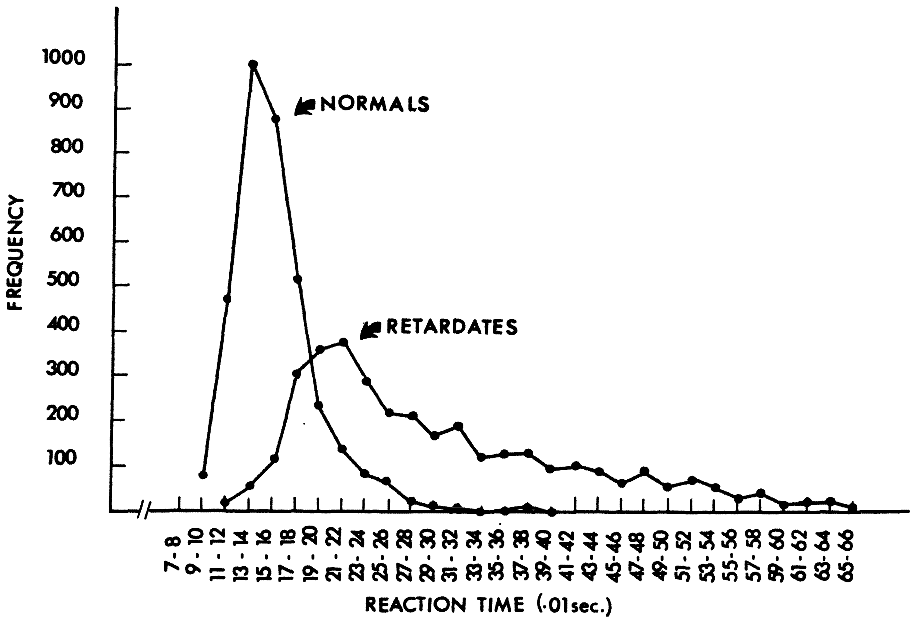


Fig. 17. Frequency distribution of 600 trials per subject of simple RT for six retarded and six normal subjects. Baumeister and Kellas (1968 b)

RT $\bar{\sigma}_i$ are positively correlated (about +.40) and the question arises as to which variable is the more fundamental aspect of IDs. It is fairly easy to imagine how IDs in RT σ_i could cause IDs in median RT, but the reverse is much harder to understand. If there were a physiological limit for the speed of RT, with negligibly small IDs in this limiting speed, and if there were considerable IDs in σ_i, then there would inevitably be considerable IDs in median RT (over *n* trails), and σ_i and median (or mean) RT would be positively correlated. IDs in *g* would be hypothesized to be related primarily to RT σ_i and the correlations of *g* with median RT, and with the intercept and slope of RT in the Hick paradigm, would all necessarily follow. One expectation from this model is that bright and dull Ss should differ not at all or only slightly in the fastest RTs of which they are capable on any trial, whereas their median RTs over *n* trails should differ considerably. A study by Baumeister and Kellas (1968 b) presents suggestive relevant data in the frequency distributions of RTs (simple RT) obtained in 600 trials for six university students and six mildly retarded (IQs 50–81, mean IQ 62), but physically normal, persons of about the same age. As shown in Fig. 17, the groups differ much less in their fastest RTs than in any measure of the central tendency of each of the two distributions. But it is also noteworthy that in a total of 3,600 trials of *simple* RT, the retarded Ss do not produce a single RT that is as fast as the 60 or 70 fastest RTs (out of 3,600) of the normal Ss. Any theory must account for this difference in the fastest possible RTs bright and retarded Ss can produce, even for simple RT. It must also account for the important fact that there is a close relationship between a S's fastest RTs and the mean or median RT over *n* trials. Liebert and Baumeister (1973) have reported correlations as high as .96 (for college students) between mean RT over 100 trials and the average of the ten fastest RTs in 100 trials. They also note that the lower limit of RT decreases with age between 6 and 18 years, as does also RT σ_i.

We have examined this phenomenon in

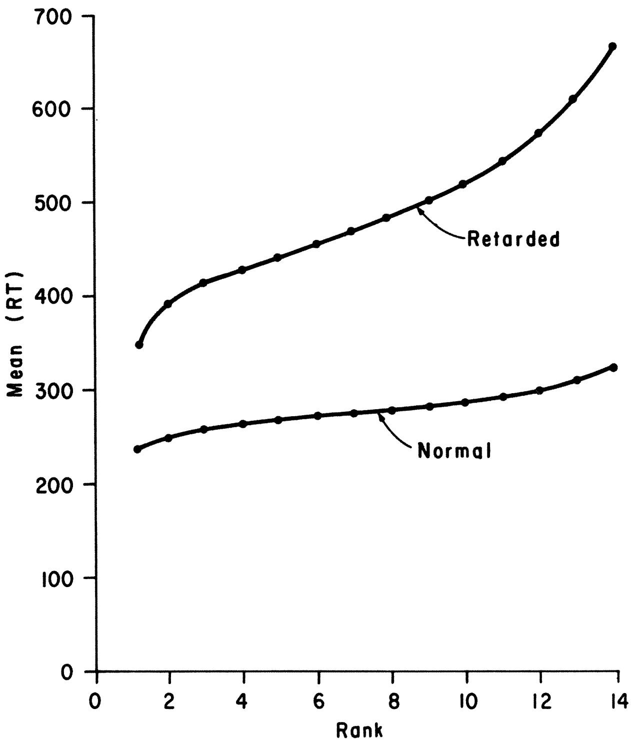


Fig. 18. Mean simple RT for 0 bits in the Hick paradigm, plotted after ranking each individual's RTs on 15 trials from the fastest to the slowest RT (omitting the 15th rank) for 46 retarded and 50 normal Ss

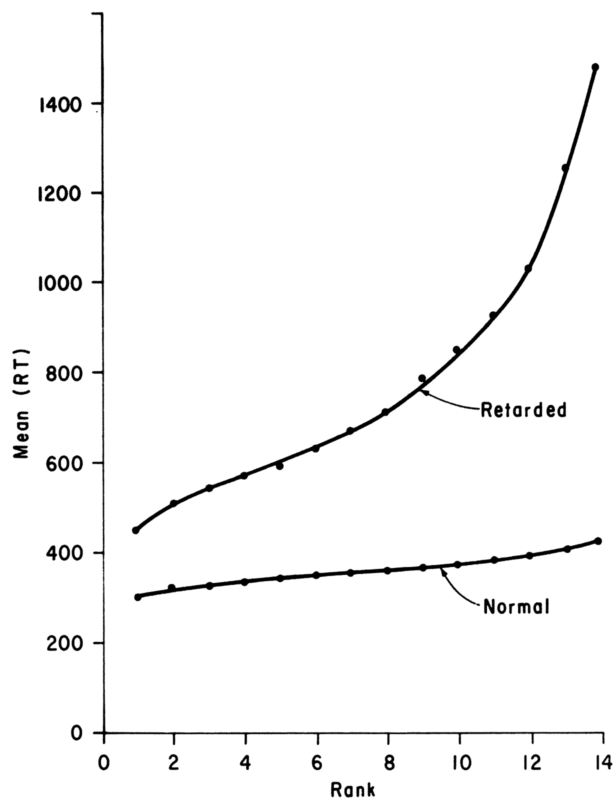


Fig. 19. Mean choice RT for 3 bits in the Hick paradigm, plotted after ranking each individual's RTs on 15 trials from the fastest to the slowest RT (omitting 15th rank) for 46 retarded and 50 normal Ss

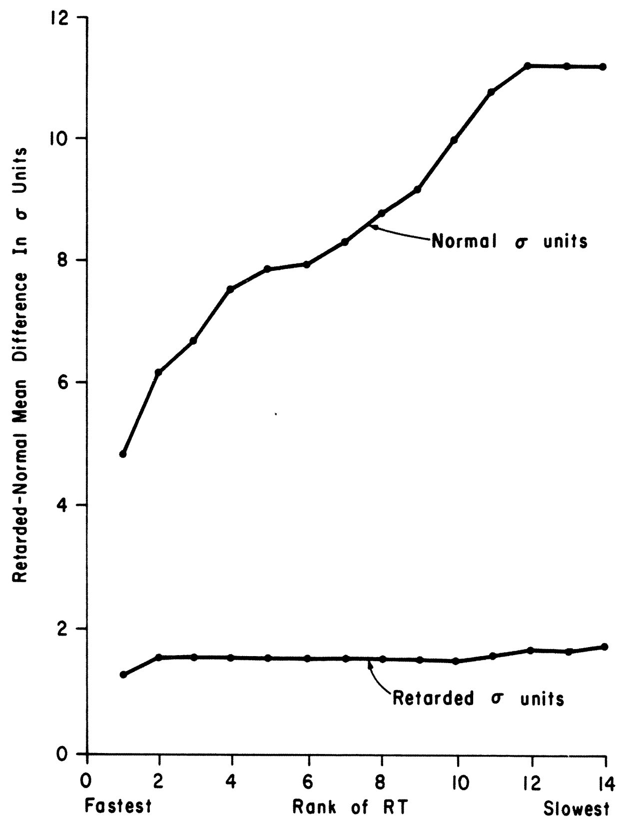


Fig. 20. The mean differences in RT between the retarded and normal groups at each rank, from fastest to slowest RTs in 15 trials, are here expressed in terms of each group's standard deviation (σ) at each rank

the Hick paradigm, using the RT−MT apparatus. Each S's RTs are rank ordered from the shortest to the longest in 15 trials. (The 15th rank is eliminated to get rid of possible outliers.) Figures 18 and 19 show the means of the ranked RTs to 0 and 3 bits, respectively, for 46 mildly retarded (mean IQ 70) and 50 bright normal young adults (mean IQ 120) given 15 trials at each level of bits. Even for simple RT, the retarded and normal groups differ by 111 ms on their fastest RT in 15 trials (rank 1); the normal group's *slowest* RT (rank 14) is 32 ms shorter than the retarded group's *fastest* RT. These differences becomes more exaggerated for choice RT involving 3 bits (i.e., eight light/button alternatives (Fig. 19), in which the fastest RTs of the retarded and normal groups differ by 142 ms.

These differences are seen to be quite substantial when viewed in terms of each group's standard deviation, i.e., in σ units, as shown for simple RT in Fig. 20. The *fastest* simple RT of the retarded and normal groups differs by 1.2 σ in terms of the retarded group's σ units and 4.8 σ in terms of the normal group's σ units.

These findings suggest that RT differences between persons who differ in *g* do not depend on complex cognitive processes, although RT differences are certainly amplified by increasing the complexity of the reaction stimulus, as can be seen in the overall difference between Fig. 18 (0 bit) and Fig. 19 (3 bits).

The S. Sternberg Short-Term Memory Scan Paradigm. This RT paradigm, invented by Saul Sternberg (1966), measures the S's speed of scanning his short-term memory for information. The S is shown a series of (usually 2–7) digits or letters (termed the "positive set") for several seconds. Then a

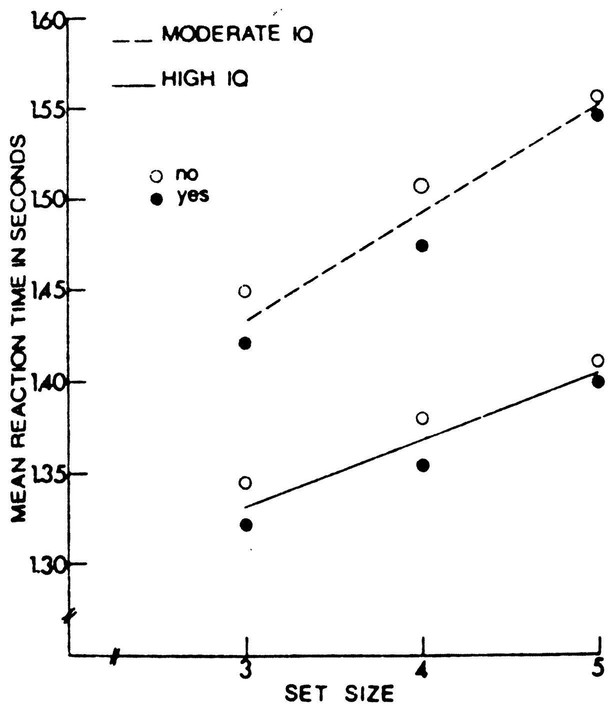


Fig. 21. Sternberg memory scan paradigm for groups of school children of moderate and high IQ, showing mean RT for determining presence ("yes") or absence ("no") of probe digit in sets of 3, 4, or 5 digits. McCauley et al. (1976)

single "probe" digit is presented. In a random half of the trials the probe digit is a member of the positive set. The S is required to respond as quickly as possible to the probe digit by pressing either a "yes" or a "no" button to indicate whether the probe was or was not a member of the positive set. RT increases linearly with size of the positive set. The fact that the ordinal position of the probe digit in the positive set has no effect on RT indicates that the scanning process is exhaustive, i.e., the S scans his memory of the entire list, regardless of where or whether the probe digit is found, although the RT is slightly longer for the absence of the probe digit than for its presence.

Several studies have shown a relationship between the intercept and slope of RT as a function of set size and mental test scores. McCauley et al. (1976), for example, applied the Sternberg paradigm to fifth and sixth grade children divided into two groups: moderate and high IQ, which yielded significantly different intercepts and slopes, as shown in Fig. 21. Keating and Bobbitt (1979) compared average and high IQ groups at ages 9, 13, and 17 years in the Sternberg paradigm, with the results shown in Fig. 22. The main effects of age,

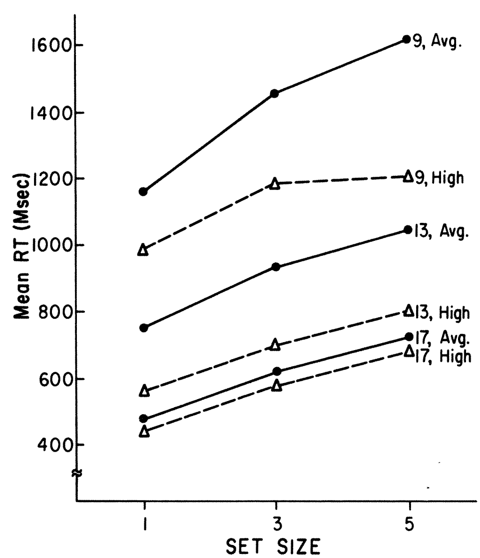


Fig. 22. Mean RT for each age/ability group as a function of number of digits in the memory set. Keating and Bobbitt (1978)

ability, and set size are all significant ($P<.001$), as is the interaction of set size and ability ($P<.05$), which accords with our generalization from the findings of simple versus choice RT and of the Hick paradigm that RT is increasingly correlated with *g* as a positive function of task complexity.

Stanford University students given the Sternberg task (Chiang and Atkinson 1976) showed much lower intercepts (about 400 ms) but showed about the same slope (i.e., a scan rate of 42 ms per digit in target set) as the high IQ children in the study by McCauley et al. (1976) (see Fig. 21) (with a scan rate of 40 ms per digit), whose IQs (with a mean of 126) are probably close to the IQs of the Stanford students. The moderate IQ group had a significantly greater slope (i.e., slower STM scanning rate) of 58 ms per digit. IQ would appear to be more crucial than mental age for short-term memory scan rate. This has interesting implications for scanning and rehearsal of information in STM to consolidate it into LTM. In terms of such a model, and in view of the observed differences in scan rates as a function of IQ, it should seem little wonder that high IQ persons in general know more about nearly everything than persons with low IQs. Snow et al. (1976) were able to "predict" the intercepts and slopes of the Sternberg memory scan paradigm for individual Stanford students, with multiple *R*'s of .88 and .70, respectively, using scores on several psychometric tests (in addition to sex). The intercept and slope parameters of the Sternberg scan, on the other hand, predicted each of four factor scores derived from a large battery of psychometric tests with *R*'s between .33 and .56. SAT-Verbal and SAT-Quantitative scores were predicted with *R*'s of .54 and .21, respectively. Remember, we are dealing here with the quite restricted range of ability in Stanford University students.

The Posner Long-Term Memory Access Paradigm. This paradigm, invented by Michael Posner (1969, Posner et al. 1969), is a measure of the time it takes a S to access a highly overlearned item of information stored in his long-term memory (LTM). The experimental procedure is based on the comparison of a S's discriminative RTs to pairs of stimuli which are the same or different either physically or semantically. For example, the letters AA are physically and semantically the same, whereas Aa are physically different but semantically the same. When Ss are instructed to respond "same" or "different" to the physical stimulus, RTs are faster than when Ss must respond to the semantic meaning. The physical discrimination is essentially the same as classical discriminative RT, but RT in the semantic discrimination involves access to semantic codes in LTM, which takes considerably more time than physical discriminative RT. The difference between semantic and physical RT thus measures access time to highly overlearned semantic codes in long-term memory.

Hunt (1976) reported the now classic experiment relating RT performance in the Posner paradigm to mental ability. Figure 23 shows these results for groups of university students who scored in the top (high) and bottom (low) quarters of the distribution of the verbal portion of the Scholastic Aptitude Test (SAT-V). AA represents the

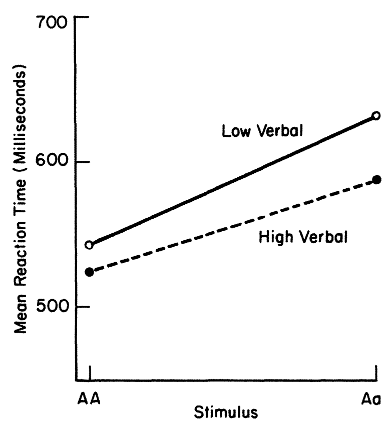


Fig. 23. Time required to recognize name identity (e.g., *Aa*) or physical identity (e.g., *AA*) of letter pairs by university students who scored in the upper (*High*) or lower (*Low*) quartile on the SAT-Verbal. Adapted from Hunt (1976, Table 1, p. 244)

physical identity choice (same-different) RT task; Aa represents the semantic identity task. University students require on the average about 75 ms more to respond to Aa than to AA types, which is the time taken by semantic encoding of the stimulus. Two features of Fig. 23 are particularly interesting in relation to findings from the Sternberg and Hick paradigms: (a) the high and low groups on SAT-V show a mean difference in RTs even on the physical, nonsemantic identity task, which is essentially just a form of classical two-choice discriminative RT; and (b) the average RT difference between AA and Aa (i.e., semantic encoding time) of 75 ms for Hunt's university students is exactly the same as the difference in RT between 0 and 3 bits of information in the Hick paradigm with university students.

Hunt's essential results with the Posner paradigm were replicated with children by Keating and Bobbitt (1978), who found significant ($P<.001$) interactions of task (physical [AA] versus semantic [Aa]) with both age and IQ level.

Unfortunately, no one has yet looked at intraindividual variability in the Sternberg and Posner paradigms or its correlation with *g*. Studies which will do so are presently underway in our laboratory.

The Nettelbeck Inspection Time Paradigm (see also Chap. 5). This method, first described by Nettelbeck and Lally (1976), measures the time required for a visual stimulus to be encoded in sufficient detail to permit a discriminative judgment. By means of a tachistoscope, the S is presented with a brief exposure of two vertical lines of markedly different length, followed by a backward masking stimulus. The S must then report whether the long line appeared on the right or the left, the position varying randomly from trial to trial. Inspection time (IT) is the duration of stimulus exposure for which the S's judgment is correct on at least 19 out of 20 trials. In highly heterogeneous groups of Ss ranging from the retarded to the gifted, correlations between IT and IQ are larger than $-.80$ (Nettelbeck and Lally 1976). Several studies that replicated this finding in small, intellectually heterogeneous groups have been reported by Brand (1979). In my laboratory, P.A. Vernon obtained a correlation of $-.31$ between IT and Raven's Advanced Progressive Matrices in a group of 25 university students – a highly restricted sample representing the top 10%–12% of high school graduates in scholastic apitude. When IT was combined with Hick paradigm RT$\bar{\sigma}_i$, the multiple R with Raven scores was .51, $P<.04$ (shrunken $R=.40$).

IT seems to reflect a very basic level of simple stimulus encoding similar to Spearman's (1927) first noegenetic law: the *apprehension* of experience. No *eduction* of relations or correlates is called for by the IT task. Yet it has shown remarkably high correlations with *g*-loaded tests in unrestricted samples. The correlation in a truly representative sample of the general population, however, remains to be determined.

Combination of Paradigms. It seems a reasonable hypothesis that these four pardigms reflect "mental speed" in each of several different systems – stimulus encoding, expectancy, scanning of short-term memory, retrieval of overlearned codes in long-term memory – and that each system contributes a unique component to IDs, in addition to a general factor in all of these variables. If this is true, and if the various cognitive systems represented by these paradigms are also operative in the much more complex information processing called for by psychometric tests, then we should expect that an optimally weighted combination of parameters derived from all four paradigms should show a much more substantial correlation with mental test scores than measurements derived from any one RT paradigm. This is exactly what Keating and Bobbitt (1978) found. Three RT-derived measures were obtained on each S: (1) choice RT minus simple RT, (2) semantic minus physical same/different RT to letter pairs (Posner paradigm), and (3) slope of RT on set size with sets of 1, 3, or 5 digits (Sternberg paradigm). The multiple R of these three measurements with Raven scores of 60 school children of average and superior IQ in grades 3, 7, and 11 was .59, .57, and .60 in the three grades, respectively. Higher correlations might be obtained if intraindividual variability were taken into account and if the correlations were corrected for attenuation, using the between-days test-retest stability coefficients. The average intercorrelation among the three paradigm measures was only .27, indicating that they are tapping different processes as well as sharing some variance in common.

If a substantial proportion of the true score variance in highly *g*-loaded psychometric tests can be "accounted for" by an optimally weighted combination of variables derived from these or other RT paradigms, it would warrant intensive investigation of the nature of IDs in these paradigms as the basis for developing an adequate theory of IDs in RT paradigms and their parameters. We are now pushing this attempt to the limit in our laboratory, using all of the previously described paradigms in combination to determine how much of the variance in psychometric *g* can be accounted for by means of these RT variables. The development of a theory of IDs in this realm, I venture, will be the essential first step toward developing a detailed theory of general intelligence. The inadequacy of the traditional and prevailing conceptions of intelligence is highlighted by the fact that they would not have predicted most of the phenomena and correlations with *g* found in the research with these several RT paradigms.

Toward a Theory of IDs in RT and *g*

Theoretical formulations of the RT phenomena I have described, and their relationship to psychometric *g*, will have to advance beyond the commonplace psychological explanations characterized by statements such as "the bright mind is the quick mind", and the like. Such generalizations, which usually are false as often as they are true, are of no help to understanding the details of the phenomena that our RT studies have revealed. Nor can we think in terms of a general "speed of work" factor which Ss bring to every kind of test or task in which they wish to excel. As I have already noted, there is often zero correlation between *g* and speed of test-taking when the test items are highly complex. Such general concepts cannot come to grips with the fine grain of the research findings relating RT to *g*, such as the correlation of *g* with intraindividual variability (σ_i), the increase in correlation between RT parameters and *g* as a function of the complexity or amount of information conveyed by the RS, Hick's law, and the systematic relationship between the σ_i of *simple* RT and the magnitude of the difference between the median of simple RT (0 bit) and two-choice RT (1 bit), and between two-choice RT and four-choice RT (2 bits), etc., in which the successive equal increments of RT as a function of information are approximately equal to the σ_i of simple RT. I believe that "easy" psychological "explanations" of these findings are suspect. If we invariably settle for an explanation of every new phenomenon in terms of a few simple and familiar psychological concepts, then the discovery and further investigation of new phenomena have no possibility of increasing our theoretical understanding of the nature of these phenomena, which virtually everyone agrees is inadequate. I also believe that adequate theoretical formulations will have to involve concepts at a molecular, neurophysiological level, rather than at just the conceptual level of psychological factors or cognitive processes.

A few well-established concepts and principles of cognitive psychology, however, afford a rationale for the importance of a time element in mental efficiency. The first such concept is that the conscious brain acts as a one-channel or *limited capacity* information processing system. It can deal simultaneously with only a very limited amount of information. The limited capacity also restricts the number of operations that can be performed simultaneously on the information that enters the system from external stimuli or from retrieval of information stored in short-term or long-term memory (STM or LTM). Speediness of mental operations is advantageous in that more operations per unit of time can be executed without overloading the system. Secondly, there is *rapid decay* of stimulus traces and information, so that there is an advantage to speediness of any operations that must be performed on the information while it is still available. Thirdly, to compensate for limited capacity and rapid decay of incoming information, the individual resorts to *rehearsal and storage* of the information into intermediate or long-term memory (LTM), which has relatively unlimited capacity. But the process of storing information in LTM itself takes time and therefore uses up channel capacity, so there is a "trade-off" between the storage and the processing of incoming information. The more complex the information and the operations required on it, the more time that is required, and consequently the greater the advantage of speediness in all the elemental processes involved. Loss of information due to overload interference and decay of traces that were inadequately encoded or rehearsed for storage or retrieval from LTM results in "breakdown" and failure to grasp all the essential relationships among the elements of a complex problem needed for its solution. Speediness of information processing, therefore, should be increasingly related to success in dealing with cognitive tasks to the extent that their information load strains the individual's limited channel capacity. The most discriminating test items thus would be those that "threaten" the information processing system at the threshold of "breakdown". In a series of items of graded complexity, this "breakdown" would occur at different points for various individuals. If individual differences in the speed of the elemental components of information processing could be measured in tasks that are so simple as to rule out "breakdown" failure, as in the several RT paradigms previously described, it should be possible to predict the individual differences in the point of "breakdown" for more complex tasks. I believe this is the basis for the observed correlations between RT variables and scores on complex *g*-loaded tests. But now we are in need of much more precise, fine-grained detail in our theoretical formulation of the phenomena than it seems cognitive theory presently has to offer.

Facts About RT with Theoretical Implications. I will here review some of the well-established findings about RT which seem to have the most suggestive implications for the development of a theory of IDs in RT and *g*. Most of these facts can be found in reviews of the effects of experimental variables on RT, such as the comprehensive chapter on RT by Woodworth and Schlosberg (1954).

1. IDs in RT are not specific to particular stimulus or response modalities. Correlations among a variety of RT procedures using different sense organs and response modes indicate that IDs in RT involve common central processes more than peripheral mechanisms. There is a substantial general factor of RT.

2. RT is related to the intensity of the reaction stimulus (RS) or the discriminability of a change in stimulation, a stronger RS producing faster RT. This suggests that the signal/noise ratio must rise above some threshold for response evocation and that increases in the signal/noise ratio (i.e., intensity of the RS) activates a greater number of the (neural) elements, increasing the probability, within a given interval of time, that the requisite threshold of neural activation will converge on the final common path for response evocation. The increase in the speed of RT as a function of RS intensity follows the Weber-Fechner law, i.e., the speed of RT increases as a linear function of the log of RS intensity. This implies a

model wherein each equal unit of increase in RS intensity activates a constant proportion of the remaining potential elements in the system that converge on the final common path, thereby monotonically increasing the probability, within a given interval of time, that the total amount of simultaneous activation will exceed the threshold for response. Increase in intensity of the RS thus makes for a negatively accelerated increase in speed of RT up to some maximum value which is limited by such factors as the activation times of sensory receptors, speed of neural conduction, muscular contractions, etc. These peripheral factors have been estimated to take up some 60–80 ms; processing in the central nervous system takes up a minimum of another 50 or 60 ms, thus making for an "irreducible minimum" RT of something between 100–150 ms. Variation in RT due to other conditions must be thought of as additions to this "irreducible minimum" of RT, hence the skewness of the distribution of RT for any individual. There are almost certainly reliable IDs in the "irreducible minimum" RT, but they are probably much smaller than IDs in the median RT under experimental conditions that add large increments to the irreducible minimum, such as an increase in the degree of uncertainty of the RS.

2. Intensity of the RS also decreases intraindividual variability (σ_i) in RT. This implies that as more elements are activated, the more "reliable" is response evocation within any interval of time. With more elements simultaneously converging on the final common path, the variance in time for reaching threshold will be reduced. If a critical number (n) of a pool of N activated elements, with random excitatory-refractory oscillations, must converge simultaneously to exceed a threshold for response evocation, the probability that n will occur within a given interval of time during which N oscillation elements are activated will increase as N increases. N is hypothesized to be a function of RS intensity.

The *area* and *duration* of a stimulus are also related to RT and RT σ_i, as both of these stimulus variables increases N, the number of activated neural elements. Because of rapid decay of the stimulus trace in the nervous system, duration of the physical stimulus becomes important by keeping N elements activated long enough for the critical N-element simultaneous activation to occur; its probability of occurrence in any interval of time decreases with a decrease in total activation, N, which falls off rapidly after the cessation of the RS. Thus, in effect, a RS of short duration is like a RS of weak intensity with respect to RT. Similarly, the *area* of stimulation affects the amount of neural activation.

These notions suggest a basis for IDs in (a) number of neural elements activated by a stimulus and (b) rate of oscillation of the excitatory-refractory phases of the activated elements. These two variables would most likely interact, because activation is transmitted throughout interconnected elements, each with a threshold of activation requiring simultaneous activation from some critical number (n) of other elements. The probability of their simultaneous convergence per unit of time would be directly related to the total number N of activated elements in the system and their rate of firing, i.e., their period of oscillation. I see oscillation as a basic concept here, not only because it is needed to help account for intraindividual trial-to-trial variability in RT, but because there are many other lines of evidence of oscillation or periodicity in the nervous system at different levels of neural organization, from refractory-excitatory oscillations in single neurones to brain waves in localized regions of the cerebral cortex involving millions of neurones, which implies a synchrony of action potential in large pools or networks of neurones. Oscillation is also a phenomenon at a chemical level; certain molecules and liquid crystals display regular rapidly oscillating structural changes over long periods. The hypothesis of IDs in the amount of hologramic neural "redundancy", i.e., the potential N of elements activated by an RS of a given intensity, area, and duration in a given sensory modality, and IDs in the rate of oscillation of activated elements (or in synchronized groups of elements) would seem to be a reasonable beginning point for the development of a theory of IDs in RT with implications for IDs in that proportion of *g* which may be shown to be correlated with RT parameters.

3. RT shows a number of interesting and theoretically suggestive parallels to phenomena in psychophysics. I have already mentioned that the relation of RT to RS intensity follows the Weber-Fechner law, which states that the increment in intensity of a stimulus necessary for a perceptible increment in sensation increases as the log of the level of stimulus intensity. Not only does the speed of RT increase as the log of RS intensity, but it decreases as the log of the number of alternatives among which the RS will occur, that is, Hick's law. There are fairly narrow boundary conditions for both the Weber-Fechner law and Hick's law, but the parallel within those conditions seems worth considering theoretically in terms of possible similar neural processes.

Just as we have found a positive correlation between IDs in simple RT (SRT) and the size of the increment in two-choice RT (CRT), the increment being median CRT minus median SRT, so, too, in psychophysics there is a positive correlation between the *absolute* threshold and the *difference* threshold, i.e., the smallest perceptible change in stimulus intensity.

As there is intraindividual trial-to-trial variability in RT, so, too, do sensory thresholds fluctuate from moment to moment. Some psychophysicists postulate an inherent Gaussian variability in thresholds and refractory periods of individual neural elements. Synchrony of individual units causes oscillation of larger groups, increasing the probability of simultaneous activation of some critical number of elements required for perceptible changes in sensation or for response evocation.

It is also interesting that momentary intraindividual *variability* in sensory discrimination is correlated with the increment in the physical stimulus needed to produce a j.n.d. (just noticeable difference) in sensation. In the Hick paradigm for RT, there is a close parallel between RT σ_i for simple RT(0 bit) and the *slope* of RT as a function of bits, i.e., the average increment in RT with each increment of information in the RS. Not only are these two variables correlated, but they are of about the same order of magnitude, as can be seen in Table 1. The correlations among the intercepts, the mean slope, and mean RT σ_i for these groups are shown in Table 2.

Table 1. Mean intercept, slope, and intraindividual variability (RT σ_i at bit) of RT in Hick paradigm for seven samples

Group	*N*	Intercept	Slope	RT σ_i at 0 bit
Mildly retarded adults	46	476.2	72.5	108.1
Elementary school children	162	305.9	39.2	42.6
Vocational college students	218	348.7	34.1	48.8
University students	25	306.4	28.4	32.3
University students	50	286.9	26.0	29.4
University students	105	305.2	30.7	32.0
University students	100	297.7	26.1	27.1
Mean of all univ. students	280	299.4	28.0	29.8

Table 2. Correlation[a] among group mean intercept, slope, and RT σ_i of the groups listed in Table 1

Variable	Intercept	Slope	RT σ_i at 0 bits
Intercept		.959	.988
Slope	.923		.987
σ_i at 0 bits	.965	.912	

[a] Above diagonal: all seven groups; below diagonal: five nonretarded adult groups

4. Choice RT increases as the physical similarity between the alternative RS increases, even when there is not the least subjective impression that changes in the degree of physical similarity of the two (or more) RS makes for any difference in their discriminability. For example, choice RS consisting of red versus yellow lights result in significantly longer choice RT than when the RS consists of red versus green lights, which are less similar than red and yellow in electromagnetic wavelengths. Presumably more similar stimulus energies produce greater overlap of excited neural elements converging on a final common path, which decreases the probability that the threshold of simultaneous activation needed for a correct discriminative reaction will be attained within a given interval of time. Greater redundancy and shorter refractory periods (i.e., faster oscillation) would increase the probability. This suggests an interesting and intuitively improbable theoretical prediction: a red-yellow choice RT task should discriminate more between high and low IQs than a red-green choice RT. (Of course there would have to be appropriate controls for stimulus intensity and color blindness, and it would be wise to use a variety of two-choice RS that differ in physical similarity). In this connection, we may recall that Spearman found that tests of pitch, brightness, and area discriminations are moderately *g* loaded (e.g. Spearman and Jones 1950, pp. 72–73, 119), and Binet included discrimination of weights as a part of his intelligence scale.

5. RT is an increasing function of the preparatory interval (PI), i.e., the interval between a "warning" or preparatory signal (PS) and the RS. This fact can be thought of in terms of the PI contributing directly to the uncertainty. Thus, even simple RT involves the uncertainty of precisely *when* the RS will occur, and Hick (1952) assumed that this uncertainty was equivalent to the increase in uncertainty resulting from one additional alternative in the number of RS. This assumption is, of course, a simplification, because we know that the amount of uncertainty as to the time of occurrence of the RS, as reflected in simple RT, varies as a function of the PI. But the fact that a PI of about 1–2 s is usually optimal for simple RT, and the fact that any shorter (or longer) PI results in longer RT, implies that there is some change in the S's "set" which facilitates RT and takes some time to attain optimal level following the PS. What, precisely, does this "preparatory set", as it is termed, consist of? A reasonable hypothesis is that it consists of a focusing or concentration (psychologically termed "attention", "alertness", or "expectancy") of the neural elements most relevant to the sensory-motor requirements of the task. Electromyograms reveal an increase in muscle tension during the PI. Also, there is a deceleration of heart rate during the PI, and mentally retarded persons show less deceleration than the nonretarded (Nettelbeck and Brewer 1981). The degree of expectancy as indicated by the increase in tension is reflected in the speed of the S's RT, although of course it is only one of a number of factors that affect RT. In *choice* RT, it seems reasonable to hypothesize, the expectancy is necessarily diffused over the two or more stimulus and response alternatives, which would reduce the redundancy of neural elements that are keyed on "ready" for any particular alternative. This might be compensated to some extent by an increase in the number of potentially activated elements involved in choice RT.

Schafer and Marcus (1973) have demonstrated a neurophysiological counterpart to expectancy, which they controlled by having Ss administer the stimulus, as contrasted to automatic presentation at random intervals, while the S's average evoked potential (AEP) to the stimulus was recorded. Self-stimulation, implying foreknowledge of the exact moment of arrival of the stimulus and hence a reduction in uncertainty, resulted in shorter latency and smaller amplitude of the AEP to both visual and auditory stimuli. The percentage reduction in amplitude under the self-stimulation condition as compared with a condition in which the subject has no control over the timing of the stimuli was termed the "self-stimulation effect". This measure, which indexes "neural adaptability", was found to be significantly related to level of intelligence, even showing a significant and striking difference between hospital technicians of average IQ and Ph.D. scientists. A subsequent larger study has further substantiated this general finding of a relationship between "neural adaptability" and psychometric intelligence. That is, people who gave larger than average evoked potentials to unexpected stimuli and smaller than average EPs to stimuli whose timing they knew as the result of self-stimulation tend to have higher IQs (Schafer 1979). A later study showed significant correlations between the "neural adaptability" measure, parameters of the RT – MT Hick paradigm, and *g* factor scores derived from a battery of 15 psychometric tests (Jensen et al. 1981). Schafer is now recording S's AEPs at the same time that Ss perform on our RT – MT apparatus. It appears that the latency of the AEP follows Hick's law, as does RT. There is undoubtedly a fairly close connection between the latencies of evoked potentials and RTs. Kutas et al. (1977) have reported correlations of +.48 and +.66 (under different conditions) between choice RT and the simultaneously recorded P300 component of the brain potential evoked by the RS. Interestingly, the P300 latencies were slower than the RT, except on the relatively few RT trials that Ss made an erroneous choice, in which case the P300 evoked potential was faster than the RT.

6. Although sense organs have analog characteristics, their output to the brain is apparently filtered through a series of "logic gates" and end up in digital form. Neurones are binary processors, i.e., they are capable of being either "on" or "off", "go" or "no go". Therefore it should not be surprising if the speed of information processing by the brain showed a binary ratio characteristic, as exemplified by Hick's law, i.e., $RT = \log_2 n$, where n is the number of alternatives of the RS. It should be noted that Hick's law is not merely peculiar to human Ss in RT experiments, but has also been demonstrated in pigeons (Blough 1977).

I have not found any attempt in the literature to explain the fact that the S Sternberg short-term memory scan paradigm yields RTs which are a linear function, not of $\log_2 n$, but simply of n, i.e., the number of *items* (not bits) in the memory set that the *S* must mentally scan. This, too, is not just peculiar to humans, but has been found to hold also for monkeys (Eddy 1973, as described by Riley 1976). My hypothesis is that the difference in outcomes between the Hick and Sternberg paradigms depends on the nature of the RS. In the Hick paradigm, the occurrence of any one of the RS alternatives immediately "rules out" all the other alternatives, and the search is ended as soon as the RS and its corresponding response are classified in this binary manner, a greater number of alternatives (n) merely taking longer as a linear function of $\log_2 n$. In the Sternberg paradigm, however, the search process (to find whether the probe digit is or is not in the memory set) requires the scanning of each single item in the series. RT data for comparable Ss in the Hick and Sternberg paradigms suggest that the same amount of time (about 30 ms for college students) is required for each *item* of the memory set in the Sternberg paradigm as is required for each *bit* of information in the Hick paradigm.

7. There is a negatively accelerated decrease in RT and in RT σ_i from early childhood up to the late teens. The form of the curve, which is a typical growth curve, is consistent with the hypothesis that some constant proportion of a limited number of undeveloped or dormant neural elements gradually becomes functional during each year of the developmental period. It is hypothesized that this growth consists of an increase in redundancy of functional neural elements, which hence increases the probability, in any unit of time, of there being simultaneously enough active elements to exceed the threshold for response. Decrease

in response latencies during the developmental period occurs in rats as well as in humans (Woodworth and Schosberg 1954, p. 36).

In humans, the decrease in RT and RT σ_i throughout the developmental period is paralleled by a decrease in the intraindividual variability of latency of the visual and auditory evoked potential (Callaway 1975, pp. 36–42).

The biological basis of these age effects is hypothesized to be the body of evidence from developmental neurophysiology which indicates that the maturing mammalian brain shows an increase in both functional capacity and the complexity of neurones (Conel 1939–1963). Although the human brain contains all the neurones it will ever have at the time of birth, the myelination of cortical nerve fibers, on which neural conductance depends, is far from complete at birth, and takes place gradually throughout the entire period of physical growth. The typical negatively accelerated growth curve would result from an approximately constant proportion of the unmyelinated neurones becoming myelinated each year.

A Model for RT in the Hick Paradigm

Hick (1952) discussed various possible types of "search" processes to find one that would best explain the phenomenon now known as Hick's law. His theoretical speculations seem obscure, which is perhaps inevitable at this stage. He stated, "With regard to the mechanism responsible for these results, speculation about neural networks is outside its present scope. There is no objection to trying to depict schematically the component operations, but it must be admitted that what analysis of the data has been carried out does little more than draw attention to the difficulties involved in finding any simple scheme" (p. 20). The model Hick proposed gave a good fit to some aspects of the data, such as the mean RT at each level of bits, but not to other aspects, such as the variances of RT at each level of bits.

The main feature of the Hick model is dictated by the necessity for hypothesizing a type of "search" process which can be thought of as successive dichotomization of the total number (n) of stimulus elements to be searched, a type of central "search" process which, on average, would take $\log_2 nt$ amount of time, where t is the time required for a single element. (This is equivalent to bits $\times t$.) I put "search" in quotes, because the RT situation does not seem to call for a search in the ordinary sense of the term. What the "search" in the Hick paradigm seems to consist of is the resoluton of uncertainty. The greater the uncertainty as to the RS, the greater the "search" (a central brain process) required for resolution, i.e., reduction of the uncertainty to zero. Why such "search" should fit a model of successive dichotomizations, each taking an equal amount of time, is not known. All that can be said at present is that this seems to be the way the nervous system operates.

Given this basic search model proposed by Hick, I have speculated about possible mechanisms that could account for the main *average* features of the RT data, as well as for IDs in these features, derived from the RT–MT apparatus. Explication of the hypothesis is facilitated by reference to Fig. 24, which depicts the dichotomizing or binary resolution of uncertainty, as measured in bits. The n choices or alternatives in the physical stimulus array can be thought of as being isomorphically represented in the neural network of the cerebral cortex. The dots in Fig. 24 represent focal points or nodes of excitation which will fire when a critical level of stimulation is reached. The number of aroused or primed nodes in the RT task corresponds to the number of alternatives in the array of RS. I hypothesize that the level of excitation at each node *oscillates*, so that half of the time the node is refractory. (The actual number of neurones involved in each node is unim-

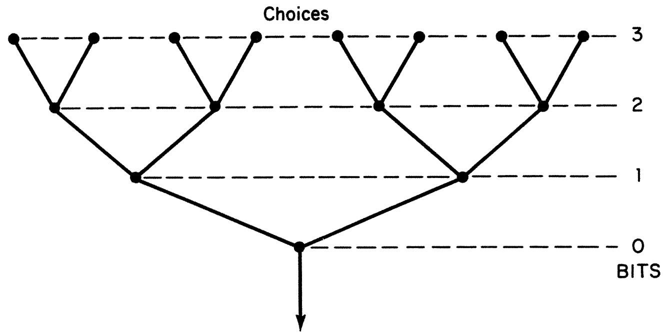


Fig. 24. Hierarchical binary tree illustrating the dichotomizing search process and the relationship of the number of choice elements to bits

portant at this point.) Above-threshold stimulation of a node at any given level (bits) is transmitted (downward in Fig. 24) through the chain of nodes to the final common path for response. For example, a RS which is one element of eight possible alternatives will excite one of the eight nodes (in top row of Fig. 24) to discharge, and the discharge will be transmitted to the final common path via three intervening nodes (at the levels of 2, 1, and 0 bits). When the RS is one of four alternatives, the excitation would be transmitted via only two intervening nodes. And so on.

The amount of time it takes to respond to the RS (over and above the irreducible minimum RT, which is attributable to peripheral sensory-motor mechanisms) hence will depend essentially on two factors: (a) the number of levels in the chain through which the excitation must be conducted, and (b) the average period of oscillation of the transmitting nodes. Excitation, of course, is not transmitted by a refractory node. Volleys of stimulation must persist until the node is excitable. The refractory phase of the oscillation at the node is the chief source of time delays in the system. IDs in the rate of oscillation would cause IDs in RT. Oscillation would also cause variability in RT from trial to trial, because the onset of the RS is random with respect to the refractory and excitatory phases of the oscillation, and we assume that the phase of oscillation of a node at any point in the chain is random with respect to the phase of any other node. Stimulation of a node at one level thus may or may not be delayed by the phase of oscillation of every other node in the chain. We have assumed for simplicity that the refractory and excitatory phases are of equal duration. The probabilities that simulation will pass through n nodes with 0, 1, 2, 3, etc. delays due to impulses arriving during refractory phases at each node conform to the binomial distribution. If p and q are the refractory and excitatory phases, respectively, and if $p=q$ and $p+q=1$, then the coefficients of the expansion of $(p+q)^n$, where n is the number of nodes in the chain, indicate the relative frequencies of there being 0, 1, 2, ..., n equal delays in a chain of n nodes. The average length of each delay will be half the time of the refractory phase of the oscillation cycle. (Speed of nerve impulses in individual neurones is so fast as to be a negligible factor in this model.) Because of the uncertainty of when the RS will occur even for simple RT, we will assume that the excitation leading to response evocation must traverse n nodes, where n is equal to bits + 1. Thus the distribution of relative frequencies of the number of delays that occur in any chain of n nodes, and the means, standard deviations, and variances of these distributions are shown Table 3. Various characteristics of these theoretical distributions can be compared with the corresponding characteristics of actual RT data obtained in the Hick paradigm using the RT−MT apparatus. It should be understood that RT is a linear function of the number of delays at the n nodes in the chain transmitting the excitation set off by the RS and leading to the response.

Table 3. Hypothetical (binomial) relative frequency distribution of time delays due to oscillation of excitatory nodes as a function of bits of information

Number of delays	Bits of information			
	0	1	2	3
0	.50	.25	.125	.0625
1	.50	.50	.375	.2500
2		.25	.375	.3750
3			.125	.2500
4				.0625
Mean	.50	1.00	1.50	2.00
σ_i	.50	.71	.87	1.00
σ_i^2	.25	.50	.75	1.00

First, note that the means in Table 3 increase as a linear function of bits, in accord with Hick's law.

Second, the σ_i of delay (=.50) at 0 bits is equal to the constant increment (=.50) in the mean delay resulting from each additional bit. (This constant increment, of course, is the slope of the Hick function.) This, too, accords with our finding that the RT σ_i at 0 bits is approximately the same absolute value as the slope of RT as a function of bits. In a samle of 280 university students, for example, the mean RT σ_i and the mean slope were 29.81 and 28.01, respectively and the correlation between IDs in these variables is almost as high as their reliabilities will permit. (Also see group comparisons in Table 1.) It is theoretically most interesting, although possibly just coincidental, that the mean critical flicker fusion (CFF) threshold in a sample of 100 of our university population is 30 Hz (i.e., 30 cycles per second), which is a light/dark cycle of 33.4 ms duration – a value remarkably close to the RT σ_i and slope of RT in this population. In terms of our bionomial oscillation model, the RT σ_i and slope of RT are equal to one-half the refractory phase of the average oscillation at a single node. It is also noteworthy that in this sample of 100 university students, there is a significant correlation ($r = +.25$, 1-tailed $P < .01$) between IDs in CFF and RT σ_i. It seems a reasonable hypothesis that the CFF threshold, that is, the rate of light/dark flicker at which subjective fusion occurs, cannot be less than about half the length of the refractory phase of the Ss rate of neural oscillation. The "neuronal filter" cannot detect a succession of stimuli as discrete if they occur at a much faster rate than the rate of neural oscillation, just as a sieve cannot screen out any mixture if the largest particles of the mixture are smaller than the sieve's finest mesh.

Third, the theoretically derived σ_i increases as a function of bits, as does intraindividual variability (RT σ_i). So far, so good. But beyond this, the simple binomial oscillation model falters. For one thing, the model's σ_i increases at a *negatively* accelerated rate as a function of bits, whereas we have found that actual RT σ_i increases in a positively accelerated fashion. A typical set of RT data, from 160 pupils in grades four to six, are plotted in Fig. 25, showing the mean RT and RT σ_i as a function of bits. The straight line and the curve are the empirically best fitting functions of the mean RT and RT σ_i, respectively.

Now, we can also fit these data to our theoretical model, which dictates that the slope of mean RTs should be equal to RT σ_i at 0 bits, and which generates RT σ_i at each level of bits. Figure 26 shows the straight line and the curve generated by the binomial oscillation model, along with the actual data points. The fit of the mean RTs to the model, of course, is very good, but the fit of RT σ_i is quite unsatisfactory – essentially the difference between a negatively (model) and positively (data) accelerated curve. In this one important specific point the binomial model fails. Could it be the case that the RT data in this particular sample ar simply anomalous with respect to RT σ_i? Before faulting the model, it would pay to look at other samples. Figures 27 and 28 show the model and data points for 218 vocational college students and 180 university students, respectively. Clearly, for both samples the discrepancy between the model-

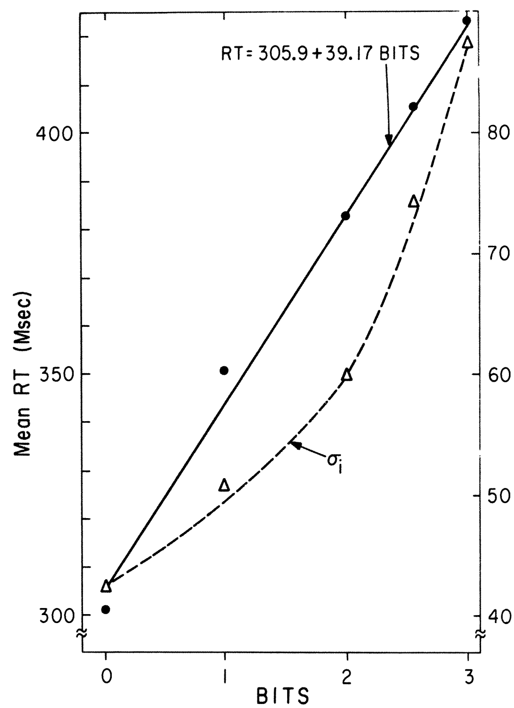


Fig. 25. Mean RT and mean σ_i of RT (i.e., mean intraindividual variability) as a function of bits in 160 pupils in grades four to six. (Note that RT and σ_i are plotted on different scales [both in milliseconds], indicated or the left and right vertical axes, respectively.)

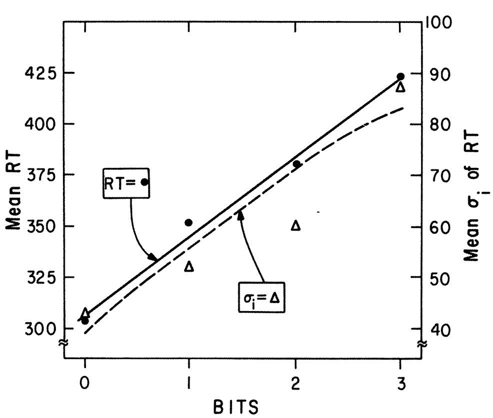


Fig. 26. The data of Fig. 25 are here fitted to the model, indicated by the straight line (predicted mean RTs) and the curve (predicted RTσ_i). Note the model's poor fit to the data points for the obtained mean RTσ_i

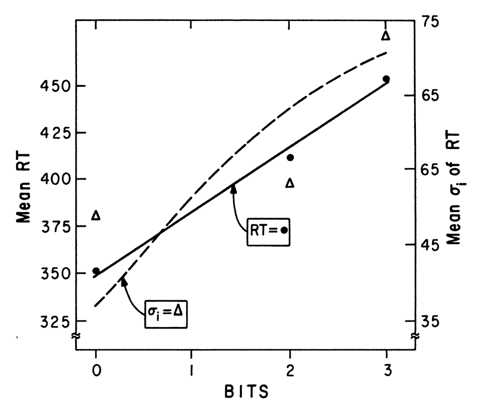


Fig. 27. RT data and predictions of mean RT and RTσ_i from binomial model (*straight line* and *curve*) for 218 vocational college students

predicted RT σ_i and the corresponding obtained values shows essentially the same difficulty as was found in the elementary school sample.

One other deficiency of this model is that it generates a symmetrical distribution of RTs at each level of bits, instead of the skewed distribution of RTs that is actually found for an individual tested on many trials. A simple but purely ad hoc improvement of the model that would produce any desired degree of skewness would be to assign unequal values to the p and q (corresponding to the relative durations of the ex-

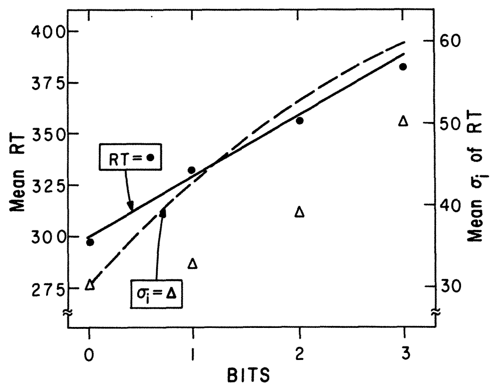


Fig. 28. RT data and predictions of mean RT and RTσ_i from the binomial model (*straight line* and *curve*) for 180 university students

citatory and refractory phases of oscillation) in the binomial equation, for example, $p = .75$ and $q = .25$. Although this ad hoc artifice will create the required skewness of the distribution of RTs, it does not cure the model's problem of negatively accelerated RT σ_i as a function of bits. Attempts so far to remedy this defect are so ad hoc as to seem unconvincing. The solution may lie in the incorporation of redundancy into the basic model, with different, and hopefully better fitting, frequency distributions being generated by *multiples* of the binary tree such as that in Fig. 24, each tree having nodes with the same frequency of oscillation, but with unsynchronized oscillations. The frequencies of refractory delays, then, would be determined by the joint action of two or more such binary trees receiving the same initial input and converging in a probabalistic fashion to exceed the excitation threshold for response. The detailed statistical implications of such a model can perhaps best be derived through computer stimulation, which we are planning to do.

Surely, the development of a mathematical-neurological model that will generate all of the specific parameters of the RT data so clearly yielded by the Hick paradigm is a priority item on the future agenda of research on the nature of IDs in RT and the mechanism of their relationship to general intelligence.

References

Baumeister AA, Kellas G (1968a) Reaction time and mental retardation. In: Ellis NR (ed) International Review of Research in Mental Retardation, vol 3. Academic Press, New York, pp 163–193

Baumeister AA, Kellas G (1968b) Distribution of reaction times of retardates and normals. Am J Ment Defic 72:715–718

Baumeister AA, Kellas G (1968c) Intrasubject response variability in relation to intelligence. J Abnorm Psychol 73:421–423

Berkson G, Baumeister AA (1967) Reaction time variability of mental defectives and normals. Am J Ment Defic 72:262–266

Blough DS (1977) Visual search in the pigeon: hunt and peck method. Science 196:1013–1914

Brand C (1981) General intelligence and mental speed: their relationship and development. In: Friedman MP, Das JP, O'Conner N (eds) Intelligence and Learning. Plenum Press, New York, pp 589–593

Callaway E (1975) Brain electrical potentials and individual psychological differences. Grune & Straton New York

Chiang A, Atkinson RC (1976) Individual differences and interrelationships among a select set of cognitive skills. Mem Cogn 4:661–672

Conel JL (ed) (1939–1963) The postnatal development of the human cerebral cortex. vol 1–4. Harvard University Press, Cambridge MA

Eddy DR (1973) Memory processes in Macaca speciosa: mental processes revealed by reaction time experiments. PhD disertation, Carnegie-Mellon University

Elliott CD, Murray DJ (1977) The measurement of speed of problem solving and its relation to children's age and ability. Br J Educ Psychol 47:50–59

Eysenck HJ (1967) Intelligence assessment: a theoretical and experimental approach. Br J Educ Psychol 37:81–98

Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47:381–391

Galton F (1883) Inquiries into human faculty and its development. Macmillan, London

Galton F (1908) Memories of my life. Methuen, London

Gilbert JA (1894) Researchers on the mental and physical development of school-children. Stud Yale Psychol Lab 2:40–100

Hick W (1952) On the rate of gain of information. Q J Exp Psychol 4:11–26

Hunt E (1976) Varieties of cognitive power. In: Resnick LB (ed) The nature of intelligence. Erlbaum, Hillsdale NJ pp 237–259

Hyman R (1953) Stimulus information as a determinant of reaction time. J Exp Psychol 45:188–196

Jensen AR (1979) g: Outmoded theory or unconquered frontier? Creat Sci Tech 2:16–29

Jensen AR (1980) Chronometric analysis of mental ability. J Soc Biol Struct 3:181–224

Jensen AR (1982) The chronometry of intelligence. In: Sternberg RJ (ed) Advances in research on Intelligence, (vol 1. Erlbaum Hillsdale NJ, pp 242–267

Jensen AR, Munro E (1974) Reaction time, movement time, and intelligence. Intelligence 3:121–126

Jensen AR, Schafer EWP, Crinella FM (1981) Reaction time, evoked brain potentials and psychometric g in the severely retarded. Intelligence 5:179–197

Keating DP, Bobbitt B (1978) Individual and developmental differences in cognitive processing components of mental ability. Child Dev 49:155–169

Kutas M, McCarthy G, Donchin E (1977) Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197:792–795

Lally M, Nettelbeck T (1977) Intelligence, reaction time, and inspection time. Am J Ment Defic 82:273–281

Lemmon VW (1927) The relation of reaction time to measure of intelligence, memory, and learning. Arch Psychol 15:5–38

Libet B (1965) Cortical activation in conscious and unconscious experience. Perspect Biol Med 9:77–86

Libet B, Alberts WW, Wright EW Jr, Feinstein B (1971) Cortical and thalamic activation in conscious sensory experience. In: neurophysiology studied in man. Procedings of a symposium, 20–22 July, 1971. Excerpta Medica, Amsterdam, pp 157–168

Liebert AM, Baumeister AA (1973) Behavioral variability among retardates, children, and college students. J Psychol 83:57–65

McCauley C, Dugas J, Kellas G, DeVellis RF (1976) Effects of serial rehearsal training on memory search. J Educ Psychol 68:474–481

Merkel J (1885) Die zeitlichen Verhältnisse der Willensthätigkeit. Philos Stud 2:73–127

Nettelbeck T, Brewer N (1981) Studies of mental retardation and timed performance. In: Ellis NR (ed) International review of research in mental retardation, vol 10. Academic Press, New York, pp 97–125

Nettelbeck T, Lally M (1976) Inspection time and measured intelligence. Br J Psychol 67:17–22

Peak H, Boring EG (1926) The factor of speed in intelligence. J Exp Psychol 9:71–94

Posner MI (1969) Abstraction and the process of recognition. In: Bower GH, Spence JT (eds) The psychology of Learning and Motivation, vol 3. Academic Press, New York, pp 43–100

Posner MI, Boies S, Eichelman W, Taylor R (1969) Retention of visual and name codes of single letters. J Exp Psychol 81:10–15

Riley DA (1976) Comments on Simon's paper, and some observations on information processing in animals. In: M.R. Rosenzweig MR, Bennett EL (eds) Neural mechanisms in learning and memory. MIT Press, Cambridge MA, pp 224–229

Roth E (1964) Die Geschwindigkeit der Verarbeitung von Information und ihr Zusammenhang mit Intelligenz. Z Exp Angew Psychol 11:616–622

Schafer EWP (1979) Cognitive neural adaptibility: a biological basis for individual differences in intelligence. Psychophysiology 16:199

Schafer EWP, Marcus MM (1973) Self-stimulation alters human sensory brain responses. Science 181:175–177

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urbana IL

Snow RE, Marshalek B, Lohman DF (1976) Correlation of selected cognitive abilities and cognitive processing parameters: an exploratory study. Technical Report 3, aptitude research project. School of Education Stanford University, Palo Alto CA

Spearman C (1927) The abilities of man. Macmillan, New York

Spearman C, Jones LLW (1950) Human Ability. Macmillan, London

Spiegel MR, Bryant ND (1978) Is speed of processing information related to intelligence and achievement? J Educ Psychol 70:904–910

Sternberg RJ (1977) Intelligence, information processing, and anological reasoning: the componential analysis of human abilities. Erlbaum, Hillsdale NJ

Sternberg S (1966) High speed scanning in human memory. Science 153:652–654

Vernon PA (1981) Reaction time and intelligence in the mentally retarded. Intelligence 5:345–355

Wade MG, Newell KM, Wallace SA (1978) Decision time and movement time as a function of response complexity in retarded persons. Am J Ment Defic 83:135–144

White PO (1973) Individual differences in speed, accuracy, and persistence: a mathematical model for problem solving. In: Eysenck HJ (ed) The Measurement of Intelligence. Williams & Wilkins, Baltimore, pp 246–260

Wissler C (1901) The correlation of mental and physical tests. Psychol Rev Monogr 3:

Woodworth RS, Schlosberg H (1954) Experimental psychology. Holt, Rinehart and Winston, New York

5 Intelligence and 'Inspection Time'

C.R. Brand and I.J. Deary

The Supposed 'Arbitrariness' of IQ Tests

In recent years it has been widely believed that IQ tests are intrinsically 'arbitrary': in so far as IQ tests reflect any real differences between people at all, these differences are said to consist merely in particular types of 'academic' ability that should properly interest only narrow-minded educational elitists; and if such differences endure through childhood this is merely because they are created and perpetuated by lasting social and educational injustices that are thought to be peculiarly prevalent under Western capitalism. Further testimony to this 'arbitrariness' of IQ tests has often been sought in the lack of any 'theoretical basis' for IQ tests. Thus the British National Union of Teachers advises its members: ".... the definition of 'intelligence' seems to rely on criteria which are subjective and social rather than objective and scientific" (Rose 1978).

Support for such views has been forthcoming from self-styled 'cognitive psychologists' – whether their primary loyalties be to traditional experimental, developmental, or differential psychology. Thus Neisser (1976) has described 'general intelligence' (*g*) as 'academic ability'; Sternberg (1980) allows merely that '.... the tests give us a reasonable measure of a fairly narrow subset of the abilities that constitute intelligence in all its manifestations'; and Hunt (1980) opines that '... the search for a "true" single information-processing function underlying intelligence is likely to be as successful as the search for the Holy Grail'. In similar reaction to difficulties that psychometricians have long acknowledged, Olson (1975) prefers to look to 'the new IQ which is based on Piaget'; and, referring to historical endeavours to find a psychological basis for conventional psychometric intelligence, Das et al. (1979) remark that 'Many a researcher has wasted his life in pursuit of a "speed" measure of intelligence'.

Some of these criticisms of general intelligence as a psychological concept show a profound disregard for: (a) the positive correlations that obtain between most measures of human ability – that is, for what Spearman (1923) called the 'positive manifold'; (b) the real theoretical insights of the founders of mental testing; and (c) the recent advances which suggest that some kind of 'mental speed' may indeed provide a psychological basis for individual differences in general intelligence (see Brand 1979, 1980). Nevertheless, it is true to say that most historical attempts to trace intelligence to mental speed have met with little more immediate support than that which greeted Sully's (1876) proposal that intelligence would turn out to be related to reaction time (RT). Even measures of solution time for IQ items themselves correlate only at around −.25 with individual IQ (e.g. Willerman 1979, Jensen 1980). Recently it has transpired that laboratory measures of choice reaction time (CRT) correlate quite substantially, at around −.45, with IQ (Jensen 1980; pp. 694,699); but it has not turned out that different CRT indices are so intercorrelated as to suggest that such mental speed is unitary (Jensen 1979a) or that technical specifications of 'number of bits of information processed per unit time' have any unequivocal relation to general intelligence (e.g. Seymour and Moir 1980). These avenues of research have shown the considerable value of Eysenck's (1967) suggestions that 'mental speed' hypotheses were worth pursuing; but they seem to lead to the conclusion that no single type of speed of 'responding', 'thinking', or 'decision-making' can account for the characteristic advantages in standard ability tests that are enjoyed by subjects of higher IQ.

All the above attempts to index mental speed, it should be noted, have been attempts to measure speed of 'output' or of hypothetical 'central processes'. However, it may be that the layman's notion of an intelligent person being 'quick on the uptake' has been unfortunately neglected by psychologists. Certainly, it is with 'average evoked potential' (AEP) measures of the brain's immediate reaction to sensory input – in the absence of any requirement for 'thought' on the part of the subject – that IQ has lately shown strong correlation (Hendrickson and Hendrickson 1980, Eysenck 1981). Perhaps Santayana had the right idea when he suggested that 'intelligence is quickness in seeing things as they really are'?

Inspection Time (IT) and IQ

The possibility of relating IQ to speed of intake of the most elementary information has existed for some time but has largely passed unnoticed. Students of basic perceptual processes have had little interest in general intelligence; and students of general intelligence have had little interest in, or knowledge of, advances in the study of perception. The requisite apparatus – the tachistoscope – has been in regular use by psychologists since the 1930's; and the capacity of a 'backward mask' to disrupt iconic storage of input has been appreciated since the 1960's (Neisser 1967). Together, what these techniques make possible is the presentation of visual stimuli extremely briefly and without the subject being able to retain any vivid image of the stimulus presentation in memory. Hence, any discrimination of the stimuli presented has to be based chiefly upon input that has only taken a small fraction of a second and which is not available to the subject in immediate memory. The stimulus duration which a subject requires to be able to make such discriminations reliably is called his inspection time (IT) (Vickers et al. 1972, Nettelbeck 1973).

As part of a programme of work with retarded people, Nettelbeck and Lally (1976) reported a study in which IT showed an astonishing correlation of $-.92$ with Performance IQ (PIQ) on the WAIS, although there was only a modest ($-.41$) correlation with Verbal IQ. Following up this lead, four studies of the relation between IT and IQ were undertaken by undergraduates in the Universities of Edinburgh and Dublin; and, by 1979, evidence was available that the relation between general intelligence and IT was both strong and robust (Brand 1979, 1981). In particular, it seemed that the IT-IQ relation, far from being restricted to Performance IQ, was more pronounced for general intelligence (and even for measures of vocabulary) than for discrete measures of spatial ability; and it appeared that the relation was not restricted to adults, but could be found even in children of only 4 years of age. At the same time, Nettelbeck's studies in Adelaide had also confirmed the PIQ-IT relation (Lally and Nettelbeck 1977, Nettelbeck et al. 1979): a review of the work in Adelaide is forthcoming (Nettelbeck and Brewer 1981). Most recently, Jensen (personal communication) has indicated that he, too, has been able to replicate the IT-IQ effect: his study of Berkeley students found a correlation which, after correction for attenuation of IQ range, would be .70 and thus well in line with the typical correlations achieved in Adelaide and Edinburgh.

The latest study of visual IT at Edinburgh is that of Deary (1980). The procedure involved is reasonably typical of the Edinburgh studies and will be presented in some detail, together with the major results.

Deary's (1980) study of the relation between IT and IQ was conducted using 13 subjects of above-average visual acuity (tested by a Snellen chart) and also adequate pitch discrimination ability – this latter ability ensured subjects would be suitable to be tested for auditory IT later (see next section). The verbal IQs of the subjects (measured by the Mill Hill Vocabulary Scale) ranged from 59 to 135, while their Raven's Progressive Matrices scores ranged from 14 to 54 (the maximum possible score is 60). Three of the subjects were hospitalized retardates without specific organic impairment; other subjects had occupations such as undergraduate, laboratory technician, policeman, nurse, milkman, domestic cleaner, and housewife.

In the IT task the subject was required to state the spatial position ('left' or 'right') of the longer of two lines presented vertically in a three-field Gerbrands tachistoscope (Model T-3 B-1). The lines were drawn with 2B pencil on white card and were 7.5 cm and 5.1 cm in length. They were 2.4 cm apart (making a visual angle of 1.6°) and were connected by a horizontal line at the top. The tachistoscope was driven from a Gerbrands six channel digital '300 series' electronic timer; field illumination was set at 90 on a Gerbrands '400 series' lamp driver. Each trial started and ended with a mask – a card containing thick matt black vertical lines designed to completely overlie the visual area that had been occupied by the stimulus and thereby prevent the further accumulation of visual information by the subject. This mask contained a central attentional dot, which served to direct subjects' gaze to the area in which the line difference would appear. Each subject was first trained to the criterion of 12 consecutive correct judgments at a stimulus duration of 130 ms (or 230 ms for the retarded subjects). Following the achievement of this training criterion subjects were given a series of decreasing exposure durations, ranging from the training criterion level to as little as 10 ms. Following this series the experimenter chose ten stimulus durations which might serve as an initial estimate of the range within which the subject's IT would lie. The range was chosen so that the subject might reasonably be expected to respond correctly for the slowest six or seven speeds but at chance level for the fastest three or four speeds.

Testing was conducted subsequently in blocks of 20 trials, i.e. a 'right' and 'left' condition for each of ten stimulus durations. The order of stimulus durations and the order of the spatial positions of the lines in any one block of trials was determined by random number tables. Each trial block was followed by a short rest pause.

After the first block had been presented the following procedure was implemented to decide whether the range of stimulus durations should be changed. If the subject had responded incorrectly in either of the two trials at any of the five slowest durations, his next block was altered: the fastest speed of the previous block was removed and a new speed, slower than the slowest speed of the previous block, was included. If the subject had responded correctly at the five slowest speeds but erred in the sixth or seventh slowest, his previous range of durations was retained. If the subject had responded correctly at all of the seven slowest speeds or more, his next block was adjusted so that the slowest speed of the previous block was removed and replaced by a duration that was faster than the fastest speed of the previous block.

The session was terminated when the subject achieved four consecutive blocks of trials whose durations, applying the above criteria, did not have to be altered. This strict criterion of response stability prevented a subject's IT being adversely affected by a poor (or lucky) start. The total testing time was approximately 45 min for normal subjects and about 1 h for retarded subjects.

The four 'stable' blocks – 80 responses in all – served as the sole data from which the IT was calculated. The IT (in milliseconds) was determined as the shortest exposure duration at which the subject was 95% correct across all durations longer than and including the level finally chosen as the IT.

The main results were as follows. Verbal IQ correlated at $-.69$ ($P<.01$) with IT; while Raven's Matrices correlated at $-.72$ ($P<.01$) with IT. These results are in line with the findings in previous Edinburgh studies and in the work of Nettelbeck and Lally.

Table 1 summarizes the nine known studies to date of the correlations between visual IT and various measures of *g*, vocabulary, and also spatial ability. The table gives details of authors, subjects, IT, and mental tests, of special features of some studies, and of the resulting correlations. There are five resulting correlations which are based on very similar studies. Each of these five independent correlations (Nettelbeck and Lally 1976, Anderson 1977, Lally and Nettelbeck 1977, Grieve 1979, Deary 1980) derives from studies involving: (a) young adult subjects – mostly male; (b) a range of IQs around 100; (c) an IT task involving comparisons of just two line lengths; and (d) non-verbal or 'culture-fair' measures of *g*. In these five, 'modal' studies, the median IT-IQ correlation is $-.80$ – which happens to be the value obtained in the single study that involved the largest number of subjects.

There are some nine major questions concerning the interpretation of these IT-IQ relations to which attention will be given in the next section. But are there any important qualifications about the general IT-IQ effect itself? (a) It is true that most of the studies have involved subjects having an artificially wide spread of IQs: standard deviations of 20 IQ points have been typical in the Edinburgh studies of young adults. Naturally, this tends to inflate the correlations between IQ and IT: correction for this spread (e.g. McNemar 1955, p. 149) would suggest that a correlation of around $-.70$ might be found if a full and representative range of normal subjects (having a standard deviation of 15 IQ points) were used. On the other hand, most measures of IQ – and certainly the measures of IT – have less than perfect internal consistency: internal reliabilities of .80 have been typical for IT in the Edinburgh studies. Such unreliability, according to classical psychometric theory, may be held to depress empirical correlations below their 'true' level (e.g. McNemar 1955, p. 159). Thus it appears that the 'true' IT-IQ correlation, even in a representative sample of the population, might be estimated as being around $-.85$. (b) Again, it would not be correct to suppose that the achieved correlations have been critically dependent upon the inclusion of mentally subnormal people in these studies. For example, Anderson's subjects of above IQ 70 showed an IT-IQ correlation of $-.64$; in Deary's study the IT-IQ correlations were $-.70$ (with Vocabulary) and $-.52$ (with Matrices) when formally subnormal subjects were omitted; and Grieve's study involved no mentally subnormal subjects at all. There is, however, some tendency (see Table 1) for IT-IQ correlations to be higher for subjects of lower and moderate levels of IQ: this will be discussed in the next section. (c) The studies in Edinburgh and Adelaide have mostly involved quite small numbers of subjects, and the confidence limits of such correlations would be wide. But such reservations must diminish when it is considered that the IT-IQ relation is apparently so robust across the many minor procedural variations that have occurred in the above studies.

In further testimony to the generality of the IT-IQ effect, there are several studies in the psychological literature which, although they do not use IT procedure or terminology, all involve tachistoscopically presented stimuli and seem to have produced compatible results.

Some of these studies involve comparisons between mentally retarded people and control subjects of much higher IQ. A review of these studies is given by Nettelbeck and Brewer (1981). But it is noteworthy that Haber and Nathanson (1969) found that the poorer recognition of serially presented words (for durations ranging from 10–150 ms) was critically related to the 'onset-to-onset' time – i.e. to the stimulus duration *plus* the time that elapsed before the onset of the next stimulus masked the first; Pennington and Luszec (1975) observed that the inferior tachistoscopic letter recognition of retardates seemed to represent a quantitative rather than a qualitative difference from normals; and the work of Maisto and Jerome (1977) suggests that temporal limitations on the stimulus may be parallelled by physical stimulus degradation – this latter is coped with more adequately by high-IQ subjects. All of the studies involving normal-retarded comparisons have used relatively complex stimuli (words, letters, digits); and it appears from them that the provision of a backward mask has not been essential to demonstrating that the retarded are handicapped with regard to information-sampling. But the use of tasks requiring more than one binary decision – and sometimes a judgment that is specified only after the stimulus is withdrawn (Mosley 1978) – seems to result in weaker effects. All these studies provide evidence for an 'intake' problem amongst low IQ subjects, but this handicap seems clearest in Nettelbeck's IT paradigm, where the subject has only to use briefly supplied information to decide between two clearly specified alternative responses.

Studies with non-retarded subjects tell a similar story: hints of IT-IQ effects can be found, but they are less strong when stimuli are comparatively complex, when they are presented for relatively long durations, and when the response alternatives are relatively numerous. Thus Tyron and Jones (1933), using 4-s presentations of prose passages, found a correlation of .57 between IQ and recognition – though 'recognition' in this study may have been partly achieved by deciphering linguistic redundancy. Livson and Krech (1956), using 200-ms presentations of dot patterns with undergraduates, found a correlation of .54 between vocabulary levels and correct identification. Most recently, Geuss (1981) reports a correlation of .28 between reasoning scores and tachistoscopic word recognition by young children.

Lastly, a different approach to the problem involves tasks that are apparently even more purely 'sensory' than IT tasks. In studies of critical flicker fusion (CFF) and dark-interval threshold (DIT) it has appeared that subjects of lower IQ and mental age are less likely to notice brief interruptions in what otherwise appears as a steady stream of light (e.g. Colgan 1954, Thor and Thor 1970). Again, the correlations with IQ are modest (around .30) and Jensen (1980, p. 710) has noted wide variability in results from CFF: such tasks might be said to involve detection of something as sensorily basic as 'change versus no change' and, in view of their long history of correlations with temperamental traits and psychopathology (Eysenck 1965, Claridge and Hume 1966) it is perhaps not surprising that they are able to reflect intelligence differences only weakly and erratically.

Thus it may be said that the IT index of mental speed is in a unique position. It appears to tap IQ differences more strongly than any other index of mental speed that has yet been tried. At the same time, it is not alone in providing testimony in favour of a mental speed theory of intelligence. It has sometimes correlated as highly as .63 with choice reaction time, for example (Lally and Nettelbeck 1977): it is arguable that IT differences, by their position at the 'intake' end of information processing, might account for many of the findings of more modest relations between such other mental speed indices and IQ: and that they might be associated with AEP complexity and thus testify that this measure is itself tapping the physiological bases of individual differences in mental speed. Together, the IT, CRT, and AEP, indices attest the correctness of Eysenck's (1981) remark: "Intelligence can now be measured by the methods of natural science – it is not just a mythical or abstract thing."

Table 1. A summary of methods and results in nine studies of the relation between visual inspection time and mental tests

Author(s)	N	Range of ages	Mental test(s)	Range of IQs	Inspection time task
Nettelbeck and	10	16–22	WAIS IQ	47–119	2 lines
Lally (1976)	10	16–22	Performance IQ	n.g.	2 lines
(Univ. Adelaide)	10	16–22	Verbal IQ	n.g.	2 lines
Anderson (1977)[a] (Univ. Edinburgh)	13	16–26	Cattell Culture-Fair (or Stanford-Binet)	44–133	2 lines
	12	16–26	Cattell Culture-Fair (or Stanford-Binet)	69–133	3 lines
	12	16–26	Cattell Culture-Fair (or Stanford-Binet)	69–133	4 lines
Lally and Nettelbeck (1977) (Univ. Adelaide)	48	17–26	WAIS Performance IQ	57–138	2 lines
Hartnoll (1978)[a] (Univ. Dublin)	18	11–12	Vocabulary + Verbal Reasoning + Verbal Fluency	'normal'[b]	Animal[c] Names
	18	11–12	Vocabulary + Verbal Reasoning + Verbal Fluency	'normal'[b]	Animal[c] Pictures
	18	11–12	Thurstone Spatial Ability	n.a.	Animal Names
	18	11–12	Thurstone Spatial Ability	n.a.	Animal Pictures
Hosie (1979)[a] (Univ. Edinburgh)	12	All 4 years	Coloured Progressive Matrices	95–125	2 lines (coloured)
Grieve (1979)[a] (Univ. Edinburgh)	10	16–28	Cattell Culture-Fair	85–122	2 lines
	10	16–28	Mill Hill Vocabulary	81–125	2 lines
	10	16–28	Revised Minnesota Paper Form Board	n.a.	2 lines
Nettelbeck et al. (1979) (Univ. Adelaide)	14	18–30	WAIS Performance IQ	*	2 lines
Deary (1980) (Univ. Edinburgh)	13	17–25	Raven's Progressive Matrices	55–125	2 lines
	13	17–25	Mill Hill Vocabulary	59–135	2 lines
Jensen (personal communication) (Univ. California)	25	*c.* 20	Advanced Progressive Matrices	*	2 lines

n.g., not given; n.a., not appropriate; n.s., not significant at $P<.1$, two-tailed; *, see columns for subjects of higher and lower IQ levels.

[a] Summarized by Brand to be (1981).

[b] Selected from a pool of 68 schoolboys: nine subjects were high in 'verbal' ability; and nine were high in 'spatial' ability.

			Correlations between IT and IQ for subjects of higher or lower IQs			
Range of ITs (ms)	Correlation between IT and IQ (or other test)	P (two-tailed)	N	Range of IQs	Correlation between IT and IQ	P
98–554	n.g.	n.g.				
98–554	$\rho = -.92$	<.01	n.g.			
98–554	$\rho = -.37$	n.s.				
15–220	$r = -.88$	<.001	6	69–97	−.98	<.001
			6	99–133	−.41	n.s.
70–200	$r = -.78$	<.01	6	69–97	−.93	<.01
			6	99–133	−.64	n.s.
90–180	$r = -.66$	<.05	6	69–97	−.87	<.05
			6	99–133	−.38	n.s.
40–284	$r = -.80$	<.001	16	57–81	−.45	<.1
			16	90–115	−.51	<.05
			16	116–130	−.17	n.s.
40–110	$\rho = -.54$	<.02	9	Lower	−.81	<.001
			9	Higher	−.31	n.s.
30–70	$\rho = +.20$	n.s.	n.a.			
40–110	$(\rho = +.08)$	n.s.	n.a.			
30–70	$(\rho = +.15)$	n.s.	n.a.			
2,000–600	$r = -.78$	<.01	8	95–113	−.65	<.1
			6	113–125	−.75	<.1
60–120	$r = -.61$	<.1	5	85–105	−.98	<.01
			5	113–125	−.45	n.s.
60–120	$r = -.88$	<.001	6	81–105	−.97	<.002
			4	110–125	+.98	=.02
60–120	$(r = -.11)$	n.s.	n.a.			
90–392	$\rho = *$	*	14	60–95	−.24	n.s.
20–700	$r = -.72$	<.01	8	55–102	−.46	n.s.
			9	105–125	−.26	n.s.
20–700	$r = -.69$	<.01	9	59–99	−.64	<.1
			8	100–135	−.65	<.1
n.g.	$r = *$	*	25	>115	−.31	n.s.

[c] Each of the five stimuli (for both names and pictures) was successively exposed for increasing durations until recognition was achieved; the subject's average S duration enabling recognition as taken as an IT. All presentations were followed by a masking stimulus.

Some Questions About the IT-IQ Relationship

Now that the broad picture of IT-IQ association has been outlined, it is proper to consider some reservations and clarifications for which empirical evidence can be adduced at this stage.

Absolute Versus Relative Levels of IT. Although psychologists themselves view any talk of an individual's 'IQ' with the reserve that comes from knowing the standard error that must surround any IQ estimate made from conventional measures, there is a further important reason why it will not be possible to talk of a person's 'IT score' except within the limits of some particular testing procedure. Simply, absolute levels of IT are variable – depending on such precise features as stimulus intensity, efficiency of the backward mask, complexity of the stimulus, difficulty of the discrimination (e.g. Vickers et al. 1972), and so on. The ITs of a range of individuals correlate very well (around .80) across experiments that themselves yield different mean levels of absolute IT (Nettelbeck personal communication); and it may conceivably turn out that IT-IQ correlations are highest for IT procedures in which even low-IQ adults need no more than 230 ms exposure (cf. Hendrickson and Hendrickson 1980). But, for the present, absolute IT values must be acknowledged to be heavily influenced by experimental parameters: for example, adult subjects of normal intelligence have lower ITs in the Edinburgh studies than in the studies conducted in Adelaide.

Backward Masking. According to the ideas developed by Nettelbeck (1973) from the work of Vickers, elimination by backward masking of variance from individual differences in short-term iconic storage should be critical to obtaining the IT-IQ relation. Yet, as indicated above, it is far from clear that this is so – especially in studies of the slower intake speeds of mentally subnormal subjects. Backward masking clearly makes IT types tasks more difficult in a way that is easy to control and specify; the Edinburgh studies may have used backward masks that were less efficient than those used in Adelaide; and Nettelbeck and Brewer (1981) suggest that backward masking particularly increases IT in subjects of lower IQ. But there are probably many ways of restricting a subject's access to the information which he requires if he is to make a decision: the stimulus can be degraded, the subject can be required to carry the information (even together with irrelevant information) in short-term memory, and so on. In so far as Nettelbeck's IT paradigm has yielded the strongest IT-IQ correlations to date, it may be because the restriction of access to evidence that is provided by backward masking is simple, is under the experimenter's control, and is compatible with subjects having a clear-cut decision which they are trying to make from the very onset of the stimulus which provides the required evidence.

Attention. It is tempting to suggest that ITs are dependent upon the attentional effort that subjects make. It is easy to believe the mentally subnormal subjects let their attention wander throughout the 40 min testing sessions that have been typical in the Edinburgh studies. Nothing could be further from the truth. Subjects of low IQ invariably find IT tasks interesting – not least because of the high degree of positive reinforcement that can be provided when the subject has so little knowledge of whether he is giving correct answers. In any case, the determination of an IT for any subject involves finding some exposure duration at which he is 95% successful: a bored or casual subject would thus not be assigned any IT at all. Again, Charman (1979) has claimed that encouraging subjects to 'pay close and critical attention' makes no difference to their performance in an IT-type task; and Lally and Nettelbeck (1980) report that the ITs of retarded subjects are not markedly influenced by encouraging them to withhold their judgments until a stimulus has been considered particularly carefully.

Task Complexity. A striking feature of present studies of RT and IQ is that little systematic attempt has been made to examine the differential importance of response complexity and stimulus complexity: these can be manipulated separately – for example, studies can be made of simple binary choices concerning aspects of complex stimuli. For IT tasks, Anderson (1977) found lower correlations at higher levels of both S and R complexity; but these different types of complexity were not distinguished. Relations between IQ and IT seem to be lower when the stimuli themselves are arguably more complex: thus Hartnoll's (1978) study found no IT-IQ relation for the identification of drawings of animals, although there was a significant IT-IQ relation when names of animals had to be recognized. Again, it generally appears that IT-IQ associations are higher when exposure times (for adults) are well below 250 ms and thus involve minimal information in the stimulus. It is perhaps when the high-IQ subject has a well-defined task that he most clearly shows what may properly be considered as his advantage at the rapid sampling of relevant information.

Speed Versus Power. In the history of psychometric intelligence, considerable importance attaches to the distinction between 'speed' and 'power' tests. Strictly speaking, this conceptual distinction is one between tests where the score depends on correct answers per unit time (such as Digit Symbol) and tests in which most subjects still fail some items no matter how long they are given (such as Vocabulary). Lest any attempt be made to associate IT abilities uniquely with conventional 'speed' tests, it must be pointed out: (a) that this conceptual distinction has no corresponding basis in any marked empirical independence of 'speed' and 'power' tests – it is widely agreed that tests of both types can provide excellent measures of *g*; and (b) that IT shows equally strong correlations with tests of both types – if anything, in the Edinburgh studies, its correlations with Vocabulary run a little higher than its correlations with time-limited administrations of tests such as Raven's Progressive Matrices and Cattell's Culture-Fair Test. The implication would be that the natural developmental history of individuals generally ensures that speed-of-intake differences not only sustain differential abilities in 'fluid intelligence', but that such initial differences serve as a foundation for the development of qualitative differences in 'crystallized intelligence'.

Subdivisions of the IQ Range. Some of the studies of visual IT in Edinburgh and Dublin have found stronger IT-IQ correlations for subjects of less than IQ 100 than for subjects above this level. In some cases these differences have been statistically significant; and Nettelbeck (personal communication) has indicated that the subjects of IQ > 115 on whom Nettelbeck and Lally (1976) reported conformed to this picture (see Table 1). Hunt (1981), too, has suggested that a similar relation might obtain between IQs and RTs in the Posner paradigm. These observations may reflect a greater unreliability of IQs around 120 than has been noted by Vernon (1979 p. 75): an analogy might be that subjects of above-average mental speed may invest their 'excess' intelligence rather diversely, thus making conventional psychometric assessment problematic. Alternatively, it may be that a relatively high level of intake speed is a sufficient but not a necessary condition of high levels of IQ – this seems to be Hunt's idea. [By contrast, some of Jensen's (1979b) observations suggest that higher levels of speed in CRT testing paradigms may be a necessary but not a sufficient condition of high IQ.] Such possibilities can hardly be adjudicated on the basis of the slight evidence that exists at present; and Deary's recent study (see Table 1) did not find any tendency for IT-IQ correlations to break down at higher levels of IQ.

Mental Age Versus Intelligence Quotient. Nettelbeck and Lally's (1976) observations give some reason to doubt whether IT improves markedly as children develop; whereas the Edinburgh studies suggest that IT is related to mental age as well as to IQ (Brand 1981). In fact neither body of data is adequately addressed to the problems: intelligence was not tested in the 7- to 10-year-old children in Nettelbeck and Lally's study; and the Edinburgh studies did not all involve the same testing apparatus. Reviewing other IT-type studies, Nettelbeck and Brewer (to be published) suggest that intake speeds may reasonably be suggested to increase up to a mental age of 10 years; and Charman (1979) provides evidence that ITs would be much slower in 60 year-old subjects than they are for subjects in their late twenties. At the other extreme of the age range, Bower (1979) has concluded that babies 'appear to have a very limited information processing rate. Many everyday events occur at a rate too high for babies to register all the relevant information'. Altogether, it seems reasonable to continue to hypothesize at this stage that intake speed improves and declines in parallel with mental age.

General Intelligence and 'Spatial Ability'. Throughout the history of psychometric studies of intelligence, a central question has always been whether 'verbal' and 'spatial' tests should be considered to tap radically different abilities. (This issue may be of some importance in the interpretation of racial differences in mental abilities – see Brand 1974.) Jensen's (1980) view is evidently that many non-verbal, Piagetian, and 'spatial' tests are highly loaded with *g*: for example, he regards Raven's Matrices as a virtually 'pure' measure of *g*. The contrary view is that distinct spatial abilities can be identified, even though they are normally hard to measure without picking up *g* variance to some extent: exposition of this thesis is provided by McGee (1979), who goes to the unusual extreme of identifying Raven's Matrices as a representative 'spatial' test. It is precisely this type of psychometric debate that should be capable of resolution if 'mental speed' were ever accepted as the vital ingredient in general intelligence. To date, one of the Edinburgh studies (Grieve 1979) has found a correlation of merely −.11 (n.s.) between IT and scores on the Revised Minnesota Paper Form Board – a classic measure of spatial ability (see Vernon 1961); while Hartnoll's (1978) study, conducted in Dublin, found a similarly non-significant relation between IT and Thurstone's measure of spatial ability. Geuss (1981) also reports no relation between tachistoscopic errors and Horn's tests of spatial ability; Lansman (1981) reports that CRTs for verbal and spatial material are uncorrelated, and that each type of CRT relates separately to psychometric measures of verbal and spatial intelligence; and Zaidel (1981) has suggested that the brain's two hemispheres, though achieving equal scores on Raven's Matrices as judged by his special testing procedures, perform differently on individual items of that test – with the left hemisphere contributing more when verbal, analytical, and, indeed, 'fluid' intelligence is required. At present, then, it seems that IT has turned out to be particularly associated with the more general forms of intelligence that typically reveal themselves readily in normal verbal development, and that spatial abilities may require their own unique story – whether or not this involves some other kind of mental speed. Of course, the 'crystallized' intelligence that is involved in vocabulary will be left intact as 'fluid' intelligence declines from age 25 ownwards: thus it may be that, beyond age 25, those non-verbal and spatial measures that still require active deployment of fluid intelligence will themselves show higher correlations with IT.

Is the IT-IQ Relation Specific to the Visual Modality? If, as argued above, mental speed is the basis of general intelligence, then it is difficult to imagine why this mental speed would be manifest solely in the visual modality. Yet, until Deary's (1980) study all investigations of the IT-IQ relation had utilized the visual mode. The main hypothesis tested by Deary (1980) was that, if they could be measured in the auditory modality, individual differences in auditory IT would also correlate with IQ. The subjects used in this part of the study were the same subjects whose visual IT-IQ correlations were given above.

After a preliminary investigation of the psychoacoustic literature and consideration of many indications and contraindications, a task was designed in which a subject heard successively, in his preferred ear, two tones of markedly different frequency; each tone was preceded and followed by white noise – intended as a masking device; the interval between the offset of tone 1 and the onset of tone 2 was approximately 1 s. The subject's task was merely to state the temporal order of the two tones, i.e. 'high – low' or 'low – high'.

The apparatus used was an in-house-constructed unit consisting of a white noise generator and stereo amplifier, allowing full control of level, balance, and tone. A pair of headphones was used to give stimulus isolation such as is achieved by the tachistoscope in the visual form of the test. A stimulus-presentation unit was constructed specially for the study. This unit is capable of producing square wave tones, of two different frequencies, for durations ranging from 1 ms up to 800 ms which can be controlled exactly by the experimenter. The unit (named a 'tachistophone') was constructed so that the white noise was exactly offset while a tone was presented. The stimuli used were two square-wave tones of 770 Hz and 880 Hz presented at a level of 90 dB. Duration of both tones was controlled by the same dial: this ensured that, when a stimulus pair was presented to the subject, the duration, intensity, and masking of each tone were identical – the only detectable difference was the frequency of the two tones. From this frequency difference alone, subjects were required to indicate the temporal order of the two stimulus tones.

Since the study was exploratory it was decided to use the same block of trials for all subjects. A block was composed of 19 durations (from 100 ms to 2.7 ms), which involved 12 trials at each duration (six had a 'high – low' temporal order and six the reverse order). This block of 228 trials was satisfactory for all subjects except one hospital patient, who managed to perform the task when the stimulus duration time was extended to include durations longer than 100 ms. All tone pairs for each duration were presented consecutively, starting with the longest durations and progressing to the shortest. The training criterion for this task was 12 consecutive correct responses at the 100 ms stimulus duration. To ensure that a subject's eventual breakdown in tone-pair discrimination was due to briefness of stimulus duration – and not fatigue – once the whole block was completed, the subject was readministered the block of 12 trials which constituted his auditory IT to check that he was still able to perform at this level. Occasional single errors made by subjects at speeds slower than their IT were ignored if the person could 'recoup' the errors by achieving 12 correct responses at a faster speed.

As in the visual IT procedure described above, an auditory IT, in milliseconds, was calculated for each subject. This was obtained by taking the shortest duration at which the subject had scored at least 11 correct responses out of 12 for *that* duration – his IT – and all slower durations. That a slightly different reckoning of auditory IT compared with visual IT was warranted should be obvious due to the larger number of trials (228) that provided the data.

The main findings were as follows. Auditory ITs in the population tested in this study ranged from 6 ms to 16 ms; and the data indicated that subjects performed virtually without errors at stimulus durations that were a few milliseconds longer than their final ITs. Verbal IQ correlated at $-.66$ ($P<.02$) with auditory IT while Raven's Matrices correlated at $-.70$ ($P<.01$) with auditory IT.

These results are certainly evidence for the hypothesis that IT indices tap a general property of the nervous system which underlies mental speed in diverse modalities. This conclusion is strengthened by the finding that auditory and visual ITs were correlated at .99 – though this correlation was dependent upon the inclusion of mentally subnormal subjects.

At Edinburgh thoughts are now being turned to the development of an auditory IT task which might improve deficiencies in the method discussed above. The temporal presentation of the tone pairs almost certainly means that the subject has to use some short-term memory in storing the first tone while waiting to receive the second – unless, of course, he is able to evaluate the absolute pitch of both tones on immediate receipt of each. To remedy this it may be possible to use a cue tone which is played for several seconds and then offset by a briefly presented comparator. The task would remain essentially the same: the subject would be required to state whether the brief tone was higher or lower than the cue tone. Secondly, white noise seemed to provide a less than perfect auditory mask – although, in the light of the uncertain evidence concerning need for a mask in the visual task, this may not be too important a reservation. A more effective mask might consist of a continuous 'warble' of both tones in the task – the warble beginning at the termination of the second tone.

Future developments being considered include the possibility of a vibrotactile IT. From an adaptation of a method used by Shiffrin et al. (1973) it would seem that an IT test could be constructed in which subjects were required to discriminate, say, temporal order (which of two fingertips was vibrated first) when vibration became very short. A spatial technique might also be used such that a subject would be required to state which finger received one of two different stimuli: these could be differences in frequency or amplitude of vibration. Backward masking of this task would consist of more vigorous vibration applied to the stimulus-receiving areas, thus preventing the accumulation of sensory information.

Implications of the IT-IQ Findings

It rightly takes 'definitions' a long while to be accepted into science: definitions, as Eysenck has pointed out, are the conclusions of scientific inquiry – not, as those people think who insist on criticizing psychometricians for 'not knowing what intelligence is', its beginning. For it to be taken as definitional of water that it boils at 100° C, it needs to be shown that occasional minor departures from that definitional truth have special explanations – in terms of impurities in the water, atmospheric pressure, and so on. A similar process of demonstration and explanation of exceptions would doubtless be necessary before the kind of 'mental speed' that is involved in inspection time could be taken as definitional to general intelligence. Indeed, even if the IT-IQ relation holds up and is not superseded by yet better indices of the relevant mental speed, there will always be a case for restricting the term 'intelligence' to what may be the typical, useful ontogenetic products of mental speed rather than letting it refer to the speed differences that naturally and normally underlie differences in reasoning, comprehension, and judgment. In any case, as was indicated in the first section, there can probably be no single definition of conventional psychometric *g* that would be acceptable to those people who prefer to maintain – for very diverse reasons – that *g* is but one interwoven feature of a rich pattern of human abilities that they are still engaged in knitting.

Let them knit on! The ten studies of IT that have either been referred to or more fully described above certainly attest the reality of *g*: as has been more fully elaborated elsewhere (Brand 1979), they militate against the view that theorizing about or working on psychometric intelligence will lead to the overthrow of IQ testing. More particularly, they constitute a sustained failure to falsify the hypothesis that intelligence has its psychological basis and developmental origin in mental speed. After 10 years in which most trait psychology has been under sustained attack (see e.g. Eysenck and Eysenck 1980) – following the collapse of behaviourism as a bastion of scientific psychology (Koch 1963) – it has turned out that IQ has correlates which will not look appealing to the champion of 'situationalist', 'labelling-theory' or traditional 'social-environmentalistic' explanations of IQ differences. These implications are perhaps quite sufficient by way of exposition of the importance of the IT-IQ relation: there is no need to insist that 'intelligence *is* mental speed', or that neither intelligence nor mental speed can ever yield to new conceptions or explanations.

Indeed, the demonstration that high-IQ subjects have a greater advantage on IT tasks than on RT tasks suggests one immediate problem for scientific research. If the high-IQ person has a speed advantage at the 'intake' stage of information processing, how does he tend to lose that advantage by the time he comes to the stage of the 'output' of information? Ruling out the hypothesis of laziness, it seems clear that processes of committing-to-memory, accessing associative memory, processing more 'deeply', reorganizing schemas, and so forth (e.g. Sternberg 1979) must now receive attention. It should be easier to study such hypothetical cognitive processes now that variance from *g* can be more securely identified. There may be other simple processes – at least as 'simple' as those of inspection time, symbol-copying, and digit span – at which the high-IQ subject performs better, even though such processes require him to 'spend' central processing time in their execution. An interesting possibility is that high-IQ subjects spend time working out exactly how they achieve correct solutions to problems (cf. Bateson 1973, part II): in this way they may 'crystallize' their intelligence so that it survives what are otherwise the ravages of aging.

The idea that mental speed – if it be the psychological basis of intelligence – requires explanation at the physiological level has been the particular concern of Hendrickson and Hendrickson (1980). Their suggestion that nervous conduction proceeds by constant-numbered pulse trains that vary between species would, if it were correct, undermine the premisses of classical neurophysiology; their stress on the role of cholinergic transmission within the CNS will surprise many physiologists, to whom other neurotransmitters have seemed more interesting in recent years; and their ideas about the function of 'engram RNA' in memory storage appear to neglect other good candidates for mediating supposed memory-transfer between individuals (Glassman 1969, Ungar et al. 1972). But were the Hendricksons correct in these matters, it would appear that they would have broken through to the physiological basis of the psychophysiological correlates of IQ which they report. It is certainly intriguing that high-IQ subjects should show more complex patterns of electrical response to simple stimuli in the two sets of data to which the Hendricksons apply their 'string-length' measure of AEP; and it is tempting to interpret these findings as suggesting that higher-IQ subjects are able to complete more successful mental work on stimulus input per unit time. In conjunction with other work showing more vigorous physiological reactions to input amongst subjects of higher IQ (e.g. for pupil dilation – see Ahern and Beatty 1980, and for palmar conductance – see Eysenck 1979), the Hendricksons' work suggests that there may be identifiable physiological bases for both 'intake speed' and 'neural efficiency'. It is to be hoped that their work will encourage other physiological psychologists to look to differential psychology – rather than to the empty promises of recent cognitive psychology – for manageable problems to which they can address their expertise.

Of course, there can be no doubt that individual differences in qualities other than intelligence itself will contribute from time to time to performance levels of individuals on particular mental tests. Thus Eysenck (1967, 1973, 1979) has long suggested that 'persistence' and 'error checking' tendencies may contribute, together with *g*, to performance on tests of fluid intelligence. To judge by the estimates that have been provided of the strength of such contributions (Eysenck 1979, Chap 8) they are relatively slight – and there is not the negative correlation between extraversion and measured intelligence that might be expected if such effects were strong. Nevertheless, it is reasonable to suggest that performance on any one intelligence test is a molecule which can be analysed to reveal the large atom of *g*.

Piagetian tests are a fascinating case of performances that deserve such analysis. Jensen (1980) insists that Piagetian items provide – as items – some of the best indices of *g* that have ever been discovered. Nevertheless, they are certainly subject to a degree of error variance; and, importantly, their tendency to correlate more highly with non-verbal, spatial tests than with vocabulary (see Eysenck 1979, Horn 1976) suggests that the *g* which they reflect is that of Jensen as opposed to the more verbally loaded *g* factor that is commonly identified in British psychometric work (e.g. Vernon 1961). In particular, the tendency of Piagetian tasks to show larger superiorities for whites versus blacks and for males versus females must arouse suspicion that they tap fluid spatial abilities (and even field independence – see Goodenough 1978) to a considerable degree. Nevertheless, Piagetian measures have sometimes been submitted to analyses which indicate that they tap information-processing efficiency of some kind (Pascual-Leone 1970, Hamilton and Launay 1976). In view of their wide popularity at present as indicators of mental development (especially in those states of the United States where IQ testing is banned by law, as has previously been the lot of IQ testing only under the most totalitarian regimes in Germany, Russia and China), it will be important to enquire whether they reflect IT or CRT variance.

Lastly, what are the practical implications of the IT-IQ relation? (a) It should be possible to test fluid intelligence in a way that is transparently fair to people of varying socio-economic, psychopathological, ethnic, national, and racial groups. Instead of testers having to say that certain individuals – autistic children, for example – are 'functioning at a level of IQ so-and-so', it should be possible to indicate mental speed levels precisely, in a way that is unaffected by educational experience, practice effects, or, indeed, lack of spatial ability. (b) IT procedures seem to allow repeated testing of subjects, which may be helpful in making assessments of clinical progress and of the effects of progressive diseases, normal aging, and alcoholism in senior personnel in positions of responsibility. [In view of the national drinking spree upon which Britain has lately embarked (Spring and Buss 1977), the latter considerations have some social importance in times of high unemployment amongst young people.] (c) It should be possible to assess ITs in subjects such as human infants and animals, for whom most existing IQ-type tests are inappropriate and with whom the vaunted techniques of cognitive psychology have made little obvious progress in recent years.

Most importantly of all, the discovery of the IT-IQ relation implies that a century of psychometry has not got one of the major features of human nature completely wrong. At a time when social science is coming under attack for not having lived up to the expectation that it would provide a new theology for post-Darwinian man, this vindication of the empirical commitments and diligent enquiries of psychometrician-psychologists is something of an achievement.

Acknowledgments. We would like to acknowledge the assistance with the studies reported in this paper of David Wight, Alan Marshall, and Tony Cull; and the contributions of Arthur Jensen, Ted Nettelbeck, Lee Willerman, and Peter Wright to our coverage and discussion of the literature.

References

Ahern S, Beatty J (1981) Physiological evidence that demand for processing capacity varies with intelligence. In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

Anderson M (1977) Mental speed and individual differences in intelligence. Final honours thesis, University of Edinburgh

Bateson G (1973) Steps to an ecology of mind. Granada, London

Bower TGR (1979) Human development. WH Freeman, San Francisco

Brand CR (1974) Intelligence differences by race and sex. London Times Higher Educ Suppl, 21st June 1974, p 18

Brand CR (1979) The quick and the educable. Bull Br Psychol Soc 32:386–389

Brand CR (1980) Myths of race and IQ. Psychol News 11:16

Brand CR (1981) General intelligence and mental speed: their relationship and development. In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

Charman DK (1979) The ageing of iconic memory and attention. Br J Soc Clin Psychol 18 2:257–258

Claridge GS, Hume WI (1966) Comparison of effects of dexamphetamine and LSD-25 on perceptual and autonomic function. Percept Mot Skills 23:456–458

Colgan CM (1954) Critical flicker frequency, age and intelligence. Am J Psychol 67:711–713

Das JP, Kirby JR, Jarman RF (1979) Simultaneous and successive cognitive processes. Academic Press, New York

Deary IJ (1980) How general is the mental speed factor in 'general' intelligence? An attempt to extend inspection time to the auditory modality. BSc (Med Sci) honours thesis, University of Edinburgh

Eysenck HJ (1965) Smoking, health and personality. Weidenfeld & Nicholson, London

Eysenck HJ (1967) Intelligence assessment: a theoretical and experimental approach. Br J Educ Psychol 37:81–98

Eysenck HJ (1973) The measurement of intelligence. MTP Press, Lancaster

Eysenck HJ (1979) The structure and measurement of intelligence. Springer, Berlin Heidelberg New York

Eysenck HJ (1981) The nature of intelligence. In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

Eysenck HJ, Eysenck MW (1980) Mischel and the concept of personality. BrJ Psychol 71:191–204

Geuss H (to be published) Information processing. 'Old wine in new bottles,' or a challenge to the psychology of learning and intelligence? In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

Glassman E (1969) The biochemistry of learning: an evaluation of the role of *RNA* and protein. Annu Rev Biochem 38:605–645

Goodenough DR (1978) Field-dependence. In: London H, Exner JE (eds) Dimensions of personality. Wiley & Sons, Chichester, pp 165–216

Grieve R (1979) Inspection time and intelligence. Final honours thesis, University of Edinburgh

Haber RN, Nathanson LS (1969) Processing of sequentially presented letters. Percep Psychophys 5:359–351

Hamilton V, Launay G (1976) The role of information-processing in the conserving operation. Br J Psychol 8:33–39

Hartnoll S (1978) Verbal and spatial ability and their relation to inspection time for word and picture stimuli. Final honours thesis, University of Dublin

Hendrickson DE, Hendrickson AE (1980) The biological basis of individual differences in intelligence. Pers Individ Diff 1:3–33

Horn JL (1976) Human abilities. Annu Rev Psychol 27:437–486

Hosie B (1979) Mental speed and intelligence: their relationship and development in four-year-old children. Final honours thesis, University of Edinburgh

Hunt E (1980) Intelligence as an information-processing concept. Br J Psychol 71:449–474

Hunt E (1981) Three faces of intelligence. In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

Jensen AR (1979a) *g*: Outmoded theory or unconquered frontier? Creat Sci Tech 2:16–29

Jensen AR (1979b) The nature of intelligence and its relation to learning. J Res Dev Educ 12:79–95

Jensen AR (1980) Bias in mental testing. Methuen, London

Koch S (1963) Psychology: a study of a science, vol 6. McGraw Hill, New York

Lally M, Nettelbeck T (1977) Intelligence, reaction time and inspection time. Am J Ment Defic 82:273–281

Lally M, Nettelbeck T (1980) Intelligence, inspection time and response strategy. Am J Ment Defic 84:553–560

Lansman M (1981) Ability factors and the speed of information processing. In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

Livson N, Krech D (1956) Dynamic systems, perceptual differentiation and intelligence. J Pers 25:46–58

Maisto AA, Jerome MA (1977) Encoding and high-speed memory scanning of retarded and nonretarded adolescents. Am J Ment Defic 82:282–286

McGee MG (1979) Human spatial abilities: psychometric studies and environmental, hormonal and neurological influences. Psychol Bull 86:889–918
McNemar Q (1955) Psychological statistics, 2nd edn Wiley, New York
Mosley JL (1978) The tachistoscopic recognition of letters under whole and partial report procedures as related to intelligence. Br J Psychol 69:101–110
Neisser U (1967) Cognitive psychology. Appleton-Century-Crofts, New York
Neisser U (1976) General, academic, and artificial intelligence. In: Resnick LB (ed) The nature of intelligence. Erlbaum, Hillsdale NJ, pp 135–144
Nettelbeck T (1973) Individual differences in noise and associated perceptual indices of performance. Perception 2:11–21
Nettelbeck T, Brewer N (1981) Studies of mild mental retardation and timed performance. In: Ellis NR (ed) Int Rev Res Ment Retard 10. Academic, New York
Nettelbeck T, Lally M (1976) Inspection time and measured intelligence. Br J Psychol 67:17–22
Nettelbeck T, Lally M (1976) Age, intelligence, and inspection time. Am J Ment Defic 83:398–401
Nettelbeck T, Cheshire F, Lally M (1979) Intelligence, work performance, and inspection time. Ergonomics 22:291–297
Olson DR (1975) The languages of experience: on natural language and formal education. Bull Br Psychol Soc 28:363–373
Pascual-Leone J (1970) A mathematical model for the transition rule in Piaget's developmental stages. Acta Psychol 32:301–345
Pennington FM, Luszec MA (1975) Some functional properties of iconic storage in retarded and non-retarded subjects. Mem Cogn 3:295–301
Rose S (1978) Race, intelligence and education: a teacher's guide to the facts and issues. National Union of Teachers, London
Seymour PHK, Moir WLN (1980) Intelligence and semantic judgment time. Br J Psychol 71:53–62
Shiffrin RM, Craig JG, Cohen E (1973) On the degree of attention and capacity limitation in tactile processing. Percept Psychophys 13:328–336
Spearman C (1923) The nature of 'intelligence' and the principles of cognition. Macmillan, London
Spring JA, Buss DH (1977) Three centuries of alcohol in the British diet. Nature 270:567–572
Sternberg R (1979) The nature of mental abilities. Am Psychol 24:214–230
Sternberg R (1980) Intelligence and test bias: art and science. Behav Brain Sci 3:35, 3:353–354
Sully J (1876) Physiological psychology in Germany. Mind 1:20–43
Thor DH, Thor CJ (1970) Dark interval threshold and intelligence. J Exp Psychol 85:207–274
Tyron CMcC, Jones HE (1933) The relation between "speed" and "altitude". J Exp Psychol 16:98–114
Ungar G, Desiderio DM, Parr W (1972) Isolation, identification and synthesis of a specific-behaviour-inducing brain peptide. Nature 238:198–210
Vernon PE (1961) The structure of human abilities. Methuen, London
Vernon PE (1979) Intelligence: heredity and environment. WH Freeman, San Francisco
Vickers D, Nettelbeck T, Willson RJ (1972) Perceptual indices of performance: the measurement of 'inspection time' and 'noise' in the visual system. Perception 1:263–295
Willerman L, Loehlin JC, Horn JM (1979) Parental problem-solving speed as a correlate of intelligence in parents and their adopted and natural children. J Educ Psychol 71:627–634
Zaidel E (1981) Hemispheric intelligence: the case of the Raven progressive matrices. In: Friedman MP, Das JP, O'Connor N (eds) Intelligence and learning. Plenum, New York

C The Psychophysiology of Intelligence

6 The Biological Basis of Intelligence. Part I: Theory

A.E. Hendrickson

Introduction: The Meaning of Intelligence

Scientific concepts originate in many cases as common semantic constructs used in our everyday languages. Thus, a child could ask his mother, 'Why do things fall to the floor when I let go of them', and be told, 'Because of the force of gravity, my dear'. Some hundreds of years ago, her answer would have been, 'Because all heavy objects fall, my dear'. The major difference between these two answers is merely semantic. If the modern child persists in asking his mother, 'What is gravity?', he will probably be told that, 'It is the force that makes heavy objects fall to the floor'. A scientist might be able to add some further facts, mention the inverse square law, and discuss the relationship of mass to acceleration. Having done so, however, the scientist cannot provide a better answer to the question, 'What is gravity?', apart from noting that it is a term used to describe a class of related phenomena.

The basic ideas that underlie our common sense notion of 'intelligence' go back to prehistory. The discovery or invention of intelligence, as a semantic concept, is probably nearly as old as the concepts of size and speed. The English language is rich in synonyms for intelligence, such as 'clever' or 'able' or 'smart' or 'lucky'. What is being discussed in this book are aspects of this common notion.

'Intelligence', then, is likewise a word which describes a class of related phenomena. We, as a race, have made some communion of observation in our everyday lives that has given rise to a group of more or less synonymous adjectives. The starting point of our theoretical investigation, then, is to rephrase the fruitless question, 'What is intelligence?', into the question, 'What observations have we made that have led us to develop this semantic group over several milleniums?'

Our discussion of this last question will attempt to introduce a measure of rigour by utilizing the relatively precise terms now available to us from the fields of information theory and engineering.

The Abstraction of Intelligence

Intelligence is a term we abstract or infer from our observation of the individual differences in some kinds of behaviour.

The behaviour that leads to our inference consists of sequences of actions which are directed towards attaining some recognizable goal. Thus, we would normally exclude motor tasks such as walking from this definition, but we might include walking as an action contained in a larger behavioural sequence. We can refer to these behavioural sequences as 'programs', which consist of sequentially related behavioural subunits called 'actions'. We will avoid a detailed definition of 'action' for the moment but note that each action (of whatever type) should take a more or less constant amount of time to perform.

It is important to note that we are not discussing individuals, but merely individual sequences or programs of behaviour.

Given that we are referring to a common goal, we can make a qualitative differentiation of behavioural programs in a number of ways.

Firstly, we might note that some programs lead to the desired goal with a higher probability of success than do other programs. Obviously, if the goal is desirable, the higher the degree of success, the better the program.

Secondly, we might note that programs differ in the amount of elapsed time from start to the attainment of the desired goal. As we have defined 'actions' as being more or less constant in time, the longer programs must contain more actions than the shorter ones. We equate a shorter elapsed time with a better program.

Thirdly, we might note that some programs require various resources to be used. If we can assign a 'cost' (energy, money, or whatever) to these resources, we can equate the lowest utilization of resources with the best programs.

Fourthly, there are the secondary considerations, or byproducts of each program. For example, one program might expose us to a certain kind of risk whilst another does not.

We will call these the probability, time, resource, and risk criteria respectively.

Using the above criteria, it should be possible to rank order a number of different programs. To do this, we have to be able to equate the criteria to a common metric, most likely by the application of certain subjective weights applied to each criterion. The individual criterion scores, times the subjective weights, are summed to a composite score.

Once we have placed the programs on a single continuum, we will want a single word to describe the attribute of ranking high or low in terms of the composite score. We can say that using a high ranking program 'is a more intelligent way to behave.' The attribute of intelligence, defined thusly, resides in the programs, and not in the individuals performing the programs.

Now we consider individuals performing programs. If we have a group of cooperative individuals, we can assign them a goal and ask them to accomplish the goal. We would normally expect to observe individual differences in their performances on one or more of the criteria.

In order to determine the origin of any differences in performance that we might observe, we need to ascertain if the same program is being used by all individuals. If different programs are in use we cannot easily say if performance differences are due to the programs or the individuals. If we can confidently say that the same program is being followed by a number of individuals, we may then infer that any observed individual differences in performance are due to some characteristic of the individuals. If we further observe that some individuals have a consistently better performance on a large variety of programs than do other individuals, we can then, and only then, attribute the quality of intelligence to the individuals as well as to programs.

In a very large class of important programs, the action subunits are not directly observable. How then can we be certain that the same programs are being used? In practice, this can only be inferred from the observation of large numbers of individuals asked to accomplish the same task under more or less identical conditions. Given certain tasks of medium difficulty, we can attempt to ensure that all individuals are aware of 'optimal' programs for the task solution, and are given training in the performance of the programs. (Difficulty is defined in terms of our criteria: a difficult task is one with a lower probability of success given a constant amount of time to attempt it, or conversely, one which requires more time to achieve the same level of probability of success as another task, or requires more resources, or exposes the individual to a higher degree of inherent risk.) We then observe the distribution of performance measures for the individuals, and relate these distributions to the degree of difficulty of the assigned tasks. If individuals did not differ except for 'errors' in teaching, training, or motivation we would expect the shape of the performance distributions to remain more or less the same as we vary task difficulty.

On the other hand, if we observe that as the task grows more difficult, there is less of a tendency for the performance distribution to skew towards the optimal performance, we are led to conclude that individuals differ in their ability to perform the various tasks. If the same individuals tend to be in the same relative place in these differing distributions, our inference is complete.

There are thus two possible sources of the concept of intelligence, corresponding to our observations of the differing qualities of behavioural programs on the one hand, and the consistent but differing quality of execution of similar quality programs by the same individuals on the other. In everyday language, we might say 'That was a smart move!', to refer to a high quality program, and 'He is a very clever chap', to describe an individual who seems to have consistent success.

These two meanings to intelligence can be referred to as 'crystallized' and 'fluid' intelligence respectively (Cattell 1963). 'Crystallized' intelligence is the greater or lesser possession of intelligent programs, and 'fluid' intelligence is the ability to execute the programs given their possession.

Individuals can be said to possess both types of intelligence. The crystallized intelligence of an individual would be some weighted sum of the programs he had in his repertoire, taking into account both the number and the quality of them.

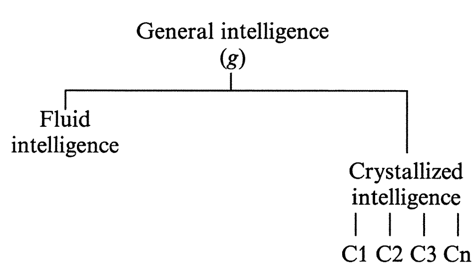
If we wish to measure the amount of fluid or crystallized intelligence resident in an individual, it is necessary to presuppose what types and kinds of program he might have stored. This presupposition must of necessity have a cultural or situational bias. It is possible to administer a set of problem goals to some individual who has few stored programs relevant to the selected goals, but who at the same time has a large set of high quality programs relevant to unselected goals.

In order to simplify the measurement task and make it as valid and universal as we can, we try to minimize the sources of sampling error. Bearing in mind the criteria for judging programs, we can attempt to hold some criteria constant across test items. This is most easily done by selecting goals that are known to have one or more programs with a perfect probability of success, require only universally held resources, and which carry no sort of risk. Presented with such goals, the individual must select the appropriate program that will accomplish the goal in the minimum amount of time.

The observation that an individual has completed a test item correctly in greater than the optimum amount of time does not in itself tell us why this has occurred. He may have selected the best program known to him, and carried it out perfectly, which would score towards high fluid intelligence and lower crystallized intelligence. On the other hand, the same elapsed time might indicate that the optimal program was selected, but executed in greater than the minimal time necessary, which would score towards high crystallized intelligence and lower fluid intelligence.

For reasons that will be discussed in detail later in this chapter, individuals coming from the same cultural backgrounds tend to be in the same place in the distribution of fluid and crystallized intelligence. We will observe that individuals high in fluid intelligence become high in crystallized intelligence as well.

As these two types of intelligence are correlated within individuals, it is convenient to speak of the composite measure of the two as a single continuum, which is referred to as general intelligence, or 'g'. This leads to the simple hierarchical diagram:



Underneath the second level of the hierarchy, we have shown a further subdivision of crystallized intelligence into a number of unnamed subfactors, labelled C1, C2, C3, ..., Cn. These correspond to the grouping of items that are found in traditional tests of intelligence. When the descriptive procedure of factor analysis is applied to a correlation matrix between individual test items, it is observed that the items will tend to group into easily interpretable subsets.

We believe that the tendency of the items to group into recognizable subsets below crystallized intelligence is due to individual differences in interest, personality, motivation, and opportunity, rather than to any fundamental differences in ability. (This does not rule out a genetic basis for these factor groupings, as interest and personality are almost certainly inherited to some extent.)

For example, two individuals, equally intelligent, might find that their interests differ substantially; one in, say, mathematics and the other in literature. It would not be surprising to find that the mathematician was the better of the two at seeing relationships in number sequences whilst the literarian had the larger vocabulary. These differences in interest and application give rise to somewhat higher correlations within common groups of test items than are observed between the various groups.

Traditional intelligence tests are subject to the criticism that they do not yield equivalent results for individuals who differ in cultural or specific ways. The two individuals referred to above would have different composite scores on intelligence tests that were weighted towards number sequences or vocabulary, respectively. The determination of the amount of 'bias' in such tests is not a question with an absolute answer, as it will largely depend on the relevance of the item content to the backgrounds of the individuals being tested.

It is fluid intelligence that is probably more interesting to most psychologists. It follows from our definitions that individual differences in fluid intelligence (if they exist at all) must have a biological basis. What biological differences between us could give rise to our more or less consistent differences in performance using the same behavioural program? Where and how are these biological differences manifested within individuals? How can we measure such differences without the 'bias' introduced by conventional intelligence measures? The rest of this chapter is devoted to an attempt to answer these questions.

The Locus of Biological Intelligence

If we assume or accept that individuals have more or less consistent differences in their ability to perform certain classes of tasks, such that we infer the existence of an underlying trait, the question of the biological locus of the individual differences in the trait is immediately posed.

In discussing what evidence we have which bears on this question, we will mention some a priori possible ways in which individuals might differ which could give rise to the performance differences we have alluded to. To facilitate this enumeration, it is useful to draw an analogy with computers. The analogy is not drawn because it is intended to imply that we humans are like computers, but is drawn primarily because the factors which underlie the performance differences between different computers are well understood.

Information Storage Capacity (Size I)

Computers differ in the amount of immediate memory (called random access memory, or RAM) they have available to them. A computer with a larger RAM can hold greater numbers of programs, and faster executing programs which are often longer than lower quality programs. It is frequently observed that there is a trade-off between the length of a program and its execution speed. For example, a simple program to count the number of bits set to '1' in a computer word (storage location) could use a simple program 'loop', testing each bit of the storage location in turn, and conditionally incrementing a sum. When the loop terminated, the sum would contain the count of the '1' bits. The execution speed of such a program would be proportional to the number of bits in the storage location, and the size of the program itself would be very small. In contrast to this, another algorithm could establish a large table in RAM which contained precomputed counts. The table would have one entry for each possible bit combination. The storage location to be counted would be used as a 'pointer' or address to retrieve the precomputed count. If our storage location had, say, 16 bits, the second program would be perhaps 30–120 times faster than the first program, but would occupy (together with the necessary table) perhaps 6,000 times as much RAM.

It is fairly evident we humans have performance ratios that are not dissimilar to the above example. One famous example of an individual with prodigious computing power and speed was the Edinburgh University professor of mathematics A. C. Aitken. Professor Aitken was capable of performing very complex calculations within seconds without any mechanical aids, which most people (including fellow mathematicians) would have required minutes or hours to do. Most people would not have been able to solve the typical 'Aitken' problem at all without the aid of pencil and paper at the very least. Aitken was aware of the basis for his manifest abilities, which lay in a large number of 'programs' and stored facts about number relationships (theorems) which are evidently not in the possession of the majority of people. Quite obvioulsy Aitken's programs had to reside somewhere in his memory. Could they exist because Aitken had more RAM than most individuals have? Or did Aitken use RAM that other individuals might have devoted to more commonplace purposes?

It seems most likely that Aitken actually had more RAM available to him than most people have. However, as will become clear in later sections of this chapter, it is believed that this is the effect of high fluid intelligence, rather than the 'cause' or explanation of it.

Instruction Primitives (Size II)

Another way in which computers differ is in the repertoire of the basic (machine level) instructions they possess. It is possible to construct a computer with a surprisingly small number of primitive instructions; yet it can be shown that such computers are so-called 'Turing machines', which are capable of computing anything a larger computer is capable of computing, given enough time.

For example, some simple minicomputers lack primitive instructions to perform the elementary arithmetic operations of multiplication and division. When the need arises to carry out a simple multiplication, this can be accomplished by a series of successive summations. This increases the amount of computer time needed for this operation by a large factor (which is data dependent) but cuts the manufacturing cost of the computer.

Note that the simple minicomputer lacking the primitive multiplication instruction needs to have a program within it to carry out the operation. No program is necessary in a more complex computer containing the primitive, apart from the single instruction needed to invoke the operation.

Carrying this example to humans, the question is posed, 'Are some individuals born with inherent mental abilities that do not exist (except as learned skills acquired later in life) in other individuals?' It seems obvious that if this situation did in fact exist, the individuals with the larger repertoire of instruction primitives would have consistent performance advantages over individuals lacking the primitives. The performance advantage would be absolute (one could, the other could not) until the deficient individuals acquired the missing skills, whereupon a relative performance advantage would exist.

If individuals did differ in terms of their instruction primitives, it raises the question of what this would imply about the biological differences we should observe which would underlie these differences. Computers that have larger numbers and more complex instructions have larger amounts and more complex circuitry. We can certainly observe the analogous differences across the phylogenetic scale; the human brain is about 1,000 times larger than a rat brain, and in some sense it must be more complex. But, do individuals within our own species differ with respect to the complexity (or quality) of the neural connections they have? Or does a high-IO person have more synapses in his brain than a low-IO person? At a lower level, could it be possible that the neurons themselves differ, such that one neuron might be more functionally capable in some internal way in a high-IO person than the corresponding neuron in a low-IO person?

There is evidence of a sort to support the notion that we are somehow 'wired' differently from each other. Some rare individuals can be observed to display mental skills of various kinds at very early ages: we might mention John Stuart Mill, Wolfgang Amadeus Mozart, and Norbert Wiener as fairly diverse examples. The environment of these famous examples was hardly ordinary, but having said that, the differences in environment cannot account for the prodigious talents of these geniuses.

What is more probable, however, is that Mill, Mozart, and Weiner had a common factor of very high '*g*', and no special 'circuitry' at all. Environment played its part in determining the use that each of these men made of their talents. Had Mozart been raised by James Mill, and Mill raised by Leopold Mozart, we probably would have found that W.A. Mozart was a famous philosopher and economist, and J.S. Mill a composer and performer of note.

Component Fabrication Technology

Computers are constructed from a variety of differing basic technologies. The very first machines used valves (vacuum tubes), which were briefly supplanted by individual transistors, which in turn have been succeeded by integrated circuits, which consist of large numbers of individual logic elements contained on single wafers, or 'chips'. The exact nature of the integrated circuits themselves differs; they vary in speed, power consumption, component density, manufacturing cost, and reliability. A computer designer can choose from a number of available technologies according to the characteristic he is trying to optimize: size, cost, speed, reliability, etc.

It is immediately evident that at the very lowest level, we humans have no differences in substances used in our fabrication. In common with all life forms, we are composed mainly of the elements carbon, oxygen, nitrogen, and hydrogen, which form the basis of most organic substances. The simple molecules, such as water or acetycholine, are not known to differ in any material way from one individual to the next.

The first level which requires some consideration is that of the complex working tools within our cells, the proteins. It is known for certain that there are individual differences here; for example, the condition known as sickle-cell anaemia is in fact the substitution of a single amino acid (valine for glutamic acid) in the beta chain of the haemoglobin molecule, which is 146 amino acid residues long. The condition (one of several known haemoglobinopathies) is known to be inherited in a Mendelian codominant fashion.

However, it does not seem likely that the alteration of one or more cell proteins could underlie the observed differences in intelligence that exist. The shape of the observed distributions of intelligence test scores (approximating to the normal curve) seems to belie this. Such a protein alteration, if one existed, would be more likely to result in two easily recognized subclasses of individual. For example, a condition such as Down's syndrome might be found to be caused by the absence or malformation of a single protein.

Far more likely is the possibility that we might differ in the amount of relative balances of the various proteins contained in our neurons. Here we can find abundant differences between individuals and even within the same individual from time to time. The brain specific protein known as S-100 for example, can be observed to be found in greater or lesser abundance throughout the lifetime of an individual.

Information Input Capacities

In addition to differences in central processing power, computers differ markedly in their capacities with respect to what is called input/output, or 'I/O' capacities.

I/O capacity can be subdivided itself. Firstly, there are the various I/O devices themselves.

One computer might have a card reader attached to it, whilst another does not. Secondly, there is an overall parameter which differentiates most computers, which is the total amount of I/O throughput that can occur in a given unit of time.

Computing tasks differ greatly in terms of the amount of raw computing power which is required relative to the amount of data which must be input. Contrast, for example, a commercial application like a payroll, which has almost trivial amounts of computing required for each data item that is input, to a multivariate statistical procedure such as factor analysis, which can require hours of pure computation following a modest amount of data input.

Obviously, we individuals do not normally differ in the numbers and kinds of our input devices; we each have two ears, two eyes, etc.

In cases where individuals have lost the function of a sense organ, or an entire sensory modality, it is doubtful if it can be said to affect their intelligence. In some senses they are less capable – a blind man has difficulty crossing busy streets. He will also find that it is more difficult to acquire abstract information, but we have little reason to believe that many kinds of problem solving ability are adversely affected. These observations and arguments would seem to rule out the possibility that observed differences in intelligence might be caused by the more invisible aspects of I/O; for example, the possibility of some damage or deficiency in the sensory pathways.

The strict analogy of computer I/O to the biological equivalent capability is difficult to discuss in a limited space, as there are rather more complex considerations of 'internal' communications channels (in both machine and man). We will pick up this point again, however, in discussing sex differences in intelligence.

Reliability

Engineers measure the reliability of a device in terms of a single parameter, which is the average amount of elapsed time between breakdowns or failures, known as mean time between failures (MTBF).

The MTBF of primitive components can be established and these individual MTBFs can be combined to give an overall MTBF of a more complex device incorporating a number of primitive components. If the probabilities of failure of a set of n individual components was the set $f1, f2, \ldots, fn$, within a given time epoch, the probability of non-failure of a device incorporating these 'n' devices would be the joint product:

$$(1-f_1)\,(1-f_2)\cdots(1-f_n)$$

and the probability of failure would be 1 minus the above joint product.

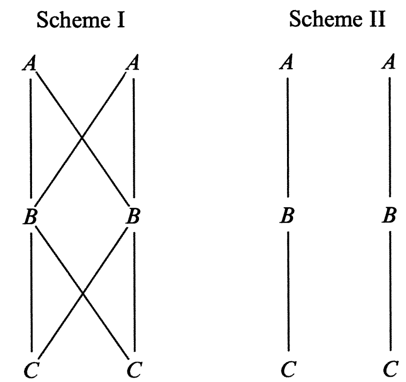
The overall MTBF is obviously affected by the individual component MTBFs. Also obvious from the above equation is the fact that the more components a device has, the more prone it will be to fail, other things being equal. An exception to this general rule, however, is when a component is added for the purpose of redundancy. Making the somewhat simplistic assumption that the interconnections between components do not themselves contribute to unreliability, if we design a device such that its overall functioning can be maintained provided at least one of a set of replicated components continues to function, a large increase in overall reliability can be established. If we have k replications of each component, then the non-failure probability of the overall device becomes:

$$(1-f_1)^k\,(1-f_2)^k \cdots (1-f_n)^k \; .$$

The obvious implication of the above equation is that it is in principle possible to achieve as much reliability as you desire. There are, however, practical problems.

The above redundancy scheme is one whereby each component is an independent entity, and it carries the implication of a large number of interconnections to affect the redundancy. The number of interconnections needed is 'k' squared times 'n', and it is therefore obvious that if the value of 'k' gets too large, the number of interconnections will be such that the overall device becomes primarily one of interconnections in terms of its mass.

An alternative redundancy scheme is to group the individual components into functional sets, such that each set has the appropriate intraconnections but no interconnections exist between sets. The number of connections needed is then simply a function of 'k', although the overall reliability is not as high. In effect, what we are doing is duplicating the entire overall entity 'k' times. Figure 1 shows the two schemes, showing a device consisting of three subcomponents, A, B, and C, and the achieved probabilities of failure associated with each scheme. We assume that each of the three components, A, B, and C, has an individual failure probability of 0.1.



Unreplicated non-failure $.9 \times .9 \times .9$	$= .729$
Unreplicated failure	.271
Scheme I non-failure $.99 \times .99 \times .99$	$= .970$
Scheme I failure	.030
Scheme II non-failure $1-(.271 \times .271)$	$= .927$
Scheme II failure	.073

Fig. 1. Two redundancy schemes, showing how three subcomponents can be interconnected to give differing levels of non-failure probability

Reliability is obviously an important characteristic of a computing system. Most of us are accustomed to thinking in terms of MTBF values of days, weeks, or even months for modern computer systems, and when a failure does occur, it is usually thought of as an inconvenience. Perhaps only the pioneers in the field will recall the MTBF values of the early computers which were seldom more than a few hours. As these early machines were also quite slow compared with their modern counterparts, the probability that one could complete a total statistical analysis before a breakdown occurred was vanishingly small. It was therefore necessary for programmers to take this fact into account, and design 're-start' points in their programs. At certain points during a programs execution, a complete copy of the computations including all intermediate values was made. When a failure occurred, the task was continued from the point of the last copy.

It is reasonably obvious that these considerations of redundancy and reliability must be of concern to us in understanding brain function and performance. The well-known ability of the brain to withstand even gross injury implies very large amounts of redundancy. Even in a normal human without any known traumas occurring to the neural-tissue, the mere fact that continuity of learned 'programs' manifestly occurs over periods of 7 decades or more (and also bearing in mind that neurons are not replaced when they die off from natural causes) inescapably implies the same thing; the brain must have massive amounts of redundancy.

What is not so obvious is the effect that these considerations might have on individual differences in intellectual performance. Consider our above example of the early computers which required special restart procedures in order to successfully complete complex statistical programs. If we were to compare such a computer with some modern hypothetical counterpart of higher reliability but the same internal speeds, we would immediately observe that each such device would execute the great majority of simple problems which were well below the respective MTBF values.

Suppose, for example, that over a 1-h time epoch the probability of successful completion of a task before failure occurred was only .10 for an old (O) technology computer, and the corresponding probability for a modern (M) technology computer the value was .90. Obviously, if one were to use these two machines on tasks that were 1 h in length, it would be very frustrating to attempt to get useful results from computer 'O'.

Now let us suppose that we give our two computers a series of tasks that are only $^1/_2$ h in length. Obviously, the probability of success for both machines is going to be higher than the 1-h failure probability given above. The $^1/_2$-h probabilities are simply the square roots of the values given above, and are .948 for computer 'M', and .316 for computer 'O'. Note that the ratio between the two $^1/_2$-h probabilities is quite different than the ratio between the 1-h probabilities.

Carrying the example a stage further, let us observe what happens when we give the two machines a 1-min task. To compute these probabilities we raise the original values to the fractional power (1/60) represented by the new time units. This gives us the values .9982 for computer M, and a value of .9623 for computer O! The absolute differences between the two machines have almost disappeared, and the ratio between them is nearly 1.0.

Consider now, the implications of these examples if it turned out that there were individual differences in the reliability of performance in mental processes. In order to demonstrate these differences, it is obvious that the length of time a given intelligence task requires is a very important factor. In our above example, a single 1-min task would have to be executed hundreds of times before we could be confident that the results were not due to chance. The $^1/_2$-h tasks would be statistically reliable with a much smaller number; as few as five would be sufficient for most purposes.

What of the biological implication of our discussion? We have shown that the reliability of a given process is affected by the reliability of the subcomponents, and is also affected by the degree and the type of redundancy organization that might be present. In the case of 'components', it may be the case that there are individual differences that affect all structural components (which we equate to individual neurons) more or less equally; some aspect of neurochemistry that can have a general and pervasive effect throughout the central nervous system. Redundancy, on the other hand, is more probably a function of environment. Learning will affect the number of replicated components and/or the interconnections between them.

We will see later that the simple notion that fluid intelligence is equated with differing levels of momentary component failure within the CNS can explain many of the known experimental facts about intelligence that we have amassed. Assuming that this model is valid, we pose the questions, where, and how does this failure occur? Indeed, what exactly do we mean by failure in this context? To answer these questions, we must first pose others of a rather fundamental sort. What is the fundamental unit of information in the CNS? How is it transmitted from one part of the brain to another? The next sections attempt to address these latter questions in detail, which we must do before returning to intelligence itself.

The Representation of Information in the CNS

Information enters the nervous system at specialized points. We can recognize a number of specialized receptor cells which are responsive to a number of physical modalities. Sound, light, smell, taste, pressure, heat and cold, and gravitational orientation are the major known modalities.

The initial response of the nervous system is to encode the physical stimuli impinging on it into nerve impulses. The sensory cells themselves seem to be dedicated to certain aspects of the stimulus qualities. Thus, for example, the cochlea is inherently able to carry out a 'spectral analysis' of sound waves by virtue of its mechanical design. Each of the 24000 or so hair cells in the human cochlea is optimally tuned to a particular frequency and can be observed to respond to a narrow range of sound frequencies.

In the case of the cochlea, the mere fact that a particular hair cell is firing is information of a sort. However, we are able to discern sound amplitude and sound patterns. These other characteristics of sound are encoded by the actual pattern of firing of each cell.

The nervous impulse, or action potential, is known to be an all or nothing type of event. A single nervous impulse (or 'pulse' for brevity) can be used to indicate such things as stimulus onset, or stimulus offset. The encoding of a continuous physical quality cannot be carried out with a single pulse, given the fact of 'all or nothing'. It is known that the cochlea uses a time-dependent scheme to encode sound amplitude. The louder the sound, the more frequently the hair cell fires. The modulation of firing rate seems to occur rapidly, and the use of frequency of firing does not seem to confound the dimension of time itself within certain limits. A pattern of rising and falling amplitude for a given frequency can in principle be represented by the firing of a single receptor cell, with the modulation of the firing rate representing the first two qualities, and the cell identity (its location along the cochlea) the third.

The action potential itself is a pseudo-electrical event. The pulse consists of a moving active zone of cell membrane through which ions are passing. The principle ionic movements are an influx of sodium (Na) into the cell, and an efflux of potassium (K) from the cell, at the leading edge of the active zone of the pulse. The ion movements reverse themselves to establish the original inside to outside ionic concentrations during the brief course of the pulse. The duration of the pulse can be divided into a short phase of ionic transfer and re-establishment lasting about a millisecond and a secondary phase lasting another 3 ms or so.

The speed of the pulse varies as a function of the diameter of the nerve axon it is travelling on. The fastest pulses travel on the thickest axons, and move at a speed of about 120 m/s. The slowest axonal transmission speeds are less than a metre per second. The frequency of firing along a single nerve axon can be up to 1000 Hz for brief periods of time, but a sustained maximum rate is of the order of 250 Hz, corresponding to the 4-ms duration of the active zone of the pulse. Stimulus-related activity along a nerve axon can be observed as a series of relatively closely spaced pulses, but it seems that there is probably a background firing rate which is not stimulus related. This background firing rate tends to be of the order of 10 Hz. We can convert these firing rates to distance: the 120 m/s speed would correspond to pulses about 0.48 metres apart at a 250-Hz firing rate. At the other end of the scale, assuming a 0.5 m/s speed, pulses would be only .002 metres apart at 250 Hz.

An interesting calculation can be made from the above figures. The mass of the human brain is approximately 1.5 litres or about 1.5×10 E15 Mm^3. If we made the rather absurd assumption that the brain was composed entirely of axons, about 1.0 Mm in diameter on average, we can compute that the total axon length would be 1.91×10 E9 m. If there were closely spaced pulses only 0.004 m apart (corresponding to the 1 Mm size), we could have about 4.77×10 E11 simultaneous pulses. At first sight this might seem a large number, but in the context of numerology of the brain, it is quite small. There are probably fewer than 6 million sensory input fibres in the human (primarily visual). If each of these fibres fired only ten times per second, and the nerve impulses persisted indefinitely, our channel capacity would be completely filled in only 7,957 s (2.2 h) given the foregoing conservative assumptions. In reality, the true capability of the CNS to represent information as a series of 'reverberating circuits' would be limited to a few seconds worth of incoming stimuli.

It is obvious from the foregoing considerations, as well as many other facts that could be cited, that the initial pseudo-electrical representation of information within the brain in the form of pulses must be converted to some other form within a short time scale. The most probable fate of the vast majority of raw information entering the nerve system is that it is encoded, 'scanned' for possible functional significance, and then lost. A certain proportion of the information must, however, be encoded and retained as a more or less permanent form of storage.

One of the most striking facts about the nervous system is the speed with which it is able to encode and decode pulses into and from the permanent storage form. For example, the memory of a familiar object, or a friend's face, or his or her voice on a telephone brings about recognition in time periods of less than a second. The familiar acts of reading and understanding of spoken language are 'real time' events with equally short translation times. Each of these events must involve the interaction of the immediately received information, represented as pulses, with the permanently held information.

Another important characteristic of the speed of interaction between pulses and permanent information storage (which henceforth we will term 'memory') is the fact that the access time seems to be fairly constant for many types of information. To illustrate this point, suppose that we unexpectedly asked somebody the name of his or her mother. In most cases, the subject of our experiment would report that the desired information 'came back' almost immediately. We could then ask the subject their date of birth, with similar results. We could add to this list almost indefinitely, with no apparent link between the questions beyond ensuring that the subject knew the answer.

The immediate implication of the foregoing consideration is that the permanent memory system within the brain must be of an associative nature. Some property or properties of the input stimuli immediately evokes the correct response. In other words, we are saying that all possible meaningful input patterns (i.e. those laid down previously as memories) are now being simultaneously matched against the incoming pulse patterns. This conclusion is reinforced when we reconsider the very slow speed of the nerve impulses. The fastest pulses travel at only 1/2000000 of the speed of electricity, and as we have seen, they cannot be spaced too close together in time. As a serial information processing device, the brain would be far too slow to act within the time scales in which it does in fact respond. We therefore conclude that the 'scanning' of incoming information for possible relevance is a parallel process, with many millions of simultaneous memory matching processes carried out for each logical input datum.

We have already noted that the encoding of pulses into 'permanent' storage must be accomplished within a short time scale. Most of the millions of bits worth of information entering the nervous system each second must be discarded, but some are recorded permanently. How do we decide what is to be encoded into permanent memory and what is not?

Consider, for a moment, the paradigm of classical conditioning. We present a subject with an arbitrary stimulus, followed closely in time by a second stimulus which innately evokes a response from the subject. After a few of these paired presentations, we can omit the second stimulus and observe that the subject will now give the response to the arbitrary stimulus by itself.

It follows from this description that the internal 'decision' to encode the arbitrary stimulus as a remembered one (which it must be to evoke the unconditioned response) can only be made after the presentation of the second stimulus. However, if the decision is made at that point, the arbitrary stimulus must still be present within the subject in some form, as a sort of 'trace'. Since we can demonstrate classical conditioning to a very broad range of initially arbitrary stimuli, it then follows that we must keep most or all incoming stimuli in some temporary form, ready to be encoded into memory if the occasion demands.

This description places constraints on the possible candidate mechanisms that we might hypothesize in an attempt to explain how the nervous system carries out these functions at a detailed level. We will now consider one such possible candidate which is believed to be in accord with these constraints.

The Neuron as an Information Processing Unit

The appearance of neurons under the light microscope varies and it is common to distinguish various types on the basis of morphology and other criteria.

In general, however, it can be stated that most neurons have a number of dendritic processes arising from the central cell body, or perikaryon, and a single axonal process. The dendritic processes serve to act as contact points primarily to the axons of other neurons or axons from primary receptor cells. The axonal process carries the outgoing pulses from the neuron to more distal neurons, bifurcating many times as it does so.

The majority of the contact points between neurons seem to be axo-dendritic synapses, although all combinations of possible synapses (axo-axonal, axo-somatic, dendrodentritic, etc.) seem to have been observed. In addition to the synapses, which are chemical transmission points, there are the so-called electrical contact points, or ephapses, which seem to be comparatively rare.

The number of synapses made by a neuron undoubtedly varies with the morphology and location of the particular cell. However, published estimates of 10000–100000 synapses for a single Purkinje cell have been made.

Stated in the most general way, the biological purpose of the neuron is to respond somehow in a logical manner to the multiplicity of information that it is receiving from moment to moment, and to send on some logical output message which is a function of the input messages. It follows from the fact that there are many logical inputs but only a single known output channel that there must either be (a) a summation or compression of the input information, or (b) highly selective responses, such that most of the input information is, in effect, ignored, but some selected bits are acted upon.

Probably because of the fact that axonal pulse transmission is carried out by the reversible passages of ions through the cell membrane, it is a popular assumption that the logical 'processing' of the arriving information, summing it in some way, must be a membrane-related process as well. It is, of course, well established that there is a postsynaptic depolarization of the (mainly dendritic) membrane with an entry of sodium ions. However, the view that these local depolarizations are somehow summed and transmitted to the axonal firing centre, as a membrane-depolarization phenomenon, is largely conjectural. There would also seem to be a number of logical criticisms that can be made of this viewpoint.

Perhaps the chief objection that can be made to the summation hypothesis is that it implies a serious loss of information. If the outgoing pulses were simply a frequency modulated quantification of the number of arriving input pulses, over time, we would have no outgoing information about which input synapses were firing. This might be acceptable in a scheme whereby the logical inputs all had an a priori meaningful relationship to each other. However, if the synaptic inputs were mainly selections from primary sensory inputs (perhaps after some 'processing' along the sensory pathways leading to the higher brain centres), what was meaningful would be the specific combinations of input stimuli that would be encountered from time to time. These combinations would not necessarily be meaningful on an a priori basis, but would become so as a result of learning, which could be viewed as a subsequent 'recognition' of a previously encountered input pattern.

Another objection to the summation hypothesis is the inability of such a scheme to account for the high degree of functional redundancy we demonstrably have. The ability of the brain to withstand a large degree of trauma is very well documented. The various experiments carried out throughout this century which sought to study brain function by systematic sectioning and ablation (e.g. K Lashley's work) demonstrated that most functions, including recently learned behavioural tasks, could be carried out by nearly decorticate animals. It is difficult to believe that such sectioning and ablation did not result in a large reduction in the numbers of synapses that were firing at any given time in some of the remaining intact cells. If so, would not this reduction have affected the degree of summation that would have been possible under such circumstances, thus rendering these remaining cells non-functional?

The above objections are given only in outline, and in addition, a number of other objections that could be made have been ignored here because of space limitations.

If the alternative hypothesis, namely, that the neuron can make selective responses to inputs, is considered, it immediately raises the question of how the information is acted upon within the neuron. If the firing of the neuron is not mediated by a consolidated membrane depolarization, what mechanism can be hypothesized in its stead?

Over the past 2 decades, the ultrastructure of the neuron has been intensively investigated by a number of means. Particularly significant are the techniques encompassed within electron microscopy and molecular neurobiology, especially in conjunction with each other.

It is now evident that within the neuron there is a rich and highly plastic molecular structure that could possibly serve to provide the required mechanism of very selective logical processing of the various neuronal inputs and the control and mediation of the outputs. This is the network of microtubules that is observed within the neuron.

Microtubules are thin hollow tubes, with a somewhat variable outer diameter, ranging from about 18 to 30 nm, with an average value being about 24 nm. The inner diameter is about 14 nm. The tubules are constructed from protofilaments, which run parallel to each other, and form a helical band. The protofiliments themselves are composed almost entirely of a globular protein subunit called tubulin. Tubulin is a dimer, with a molecular weight of about 110000. The two monomers seem to be equal in size; i.e. 55000 daltons.

Microtubules are now known to be constituents of most cells, and therefore their role and function within the neuron must be a specialist one. It is possible, of course, that microtubules have a number of functions within some cells, and other roles of a general sort would not necessarily be inconsistent with a specialist function within neurons.

What evidence is there that microtubules might have a special communications and/

or logical processing function in brain? A summary is as follows:

1. Amount. A crude indication of the importance, if not the actual function, is the sheer amount of tubulin in brain. Tubulin has been estimated to account for 15%–40% of the soluble protein of various brain extracts (Shelanski 1973).
2. Plasticity. Microtubules are seen to be highly volatile objects, which can rapidly form and dissociate. The half-life of the tubulin subunits is quite short, ranging from 4 to 5 days (Shelanski et al. 1972).
3. Self-Assembly. This seems to be a most important characteristic of the tubulin subunit; that is, no special enzyme is needed to cause microtubules to form from the basic tubulin subunits! The main requirements for assembly in vitro were the presence of a nucleotide triphosphate (GTP or ATP) and the $Mg\ 2+$ ion (Scheele and Borisy 1979).
4. Interaction with a common synaptic ion (Calcium). Calcium concentrations can control both assembly and disassembly of microtubules. Very low concentrations of $Ca\ 2+$ stimulate assembly of microtubules and high concentrations can prevent polymerization and actually disrupt pre-assembled microtubules in vivo (Luduena 1979). It is also known that the calcium regulatory protein (calmodulin, or CDR) may mediate the interaction between tubulin and calcium. CDR is known to be present at both pre- and postsynaptic sites.
5. Direct medical/biochemical implications. The severe memory disorder known as Alzheimer's disease is characterized by neurofibrillary changes that can be observed through the light microscope. Grundke-Iqbal et al. (1979) have shown by the use of immunolabelling that the '... neurofibrillary tangles in Alzheimer's disease probably originate from neurotubules'. This lends direct support to the notion that microtubules play a role in the memory function.
6. Locus. If microtubules are to have the function ascribed to them by the present hypothesis, there should be evidence that they are located at the appropriate places within the cell. There is, in fact, abundant evidence that microtubules form a coherent pattern running from the dendritic synapses down to the beginning of the axon. The most important locations are considered in detail:
   a) The postsynaptic density. This structure is a dark staining area which appears as a thickening of the membrane on the distal side of the synaptic cleft. Walters and Matus (1975) have established that '... tubulin (is a) major component of the postsynaptic density, where it probably has an important structural role in providing a matrix for more specialised proteins of functional importance in synaptic transmission'. They go on to note that 'Some form of functional role for the tubulin itself cannot, however, be discounted at present'. Microtubules can often be observed in micrographs near the postsynaptic density, where they are usually seen in cross section. A particularly good micrograph is to be seen in a study by Taxi (1967).
   b) Dendrites. According to the classic survey of Peters et al. (1970) 'Microtubules are the most prominent elements in the cytoplasm of large dendrites arising from multipolar cells. They are usually numerous and they funnel into the base of the dendrite to become arranged parallel to one another. In transverse sections ... the microtubules appear to be disposed in an array of almost crystalline orderliness'.
   c) The cell body (Perikaryon). Again, in the survey of Peters et al. (1970)

they note 'Two ... conspicuous elements in the cytoplasm of the perikaryon are microtubules and neurofilaments ... In the neuronal perikaryon, neurofilaments and microtubules occupy most of the space not pre-empted by the larger and more prominent organelles ... At lower magnifications it is apparent that, although their orientation is not completely orderly, these (microtubules) tend to run parallel to one another in loose bundles that course, like the traffic in city streets, around the more agglomerated organelles, ...'

d) The axon hillock and the initial axon segment. The nerve impulse originates in this region of the neuron, and it is therefore of some interest to note the observations of Peters et al. (1970) that '... at the base of the axon hillock many of the microtubules change their orientation so that they funnel into it. As the axon hillock becomes narrower, some of the microtubules come together to form fascicles .... These fascicles then pass into the initial axon segment, of which they form a very characteristic and diagnostic feature. ... In transverse sections ... it can be seen that the microtubules in a fascicle are separated from one another by a distance of about 250 Å. Furthermore, the members of each fascicle are connected by cross bands that pass between the microtubules like the rungs of a ladder'.

e) The axon beyond the initial segment. A few microtubules continue down into the axon proper. Peters et al. do not report any cross bridges between the microtubules that are seen to continue into the axon beyond the initial segment. They do note that 'In contrast to large dendrites, large axons contain relatively few microtubules and many neurofilaments, both of which are oriented parallel to the long axis of the process'.

f) Presynaptic axon terminals. Peters et al. (1970) note that '... microtubules, ... also continue into preterminal axons.... At boutons terminaux the microtubules seem to end short of the terminal enlargement'. Microtubules are also known to make close associations with synaptic vesicles, and may actually connect to them. Jarlfors and Smith (1969) found '... clusters of four to six vesicles (occurring) around single neurotubules'. They also noted that "... (the) association does not extend to the immediate focus of the synapse, where vesicles are closely but randomly packed'.

g) Associated structures. In order to carry out a coordination role and/or a logical processing one, it is necessary that there be some means of communicating between adjacent microtubules. There are, in fact, easily observed so-called cross-bridge structures that can be seen in many micrographs, which appear to be functional connections between microtubules (Hyams and Stebbings 1979, p. 508).

The above evidence adds up to a remarkable picture when it is taken together. Microtubules seem to form an organized network, running from dendritic synapses down to the firing centre of the neuron, at the axon hillock, where they actually communicate with each other by means of cross bridges. It seems likely that the microtubules act to carry ionic concentrations from the synapses to the first axon segment. The greatly restricted volume within the tubules should allow a rapid transport of a ionic charge, somewhat like a 'billiard ball' principle, whereby the entry of the ion at one end of the tubule is rapidly promulgated as an increased concentration throughout, thereby causing a quick egress of ions at the distal end.

It is obvious that this picture offers us a possible means whereby the neuron could make highly selective responses to the pattern (as opposed to simply the sum) of the incoming pulses. We merely need to suppose that the neuronal firing centre has a number of sites which are associated with particular firing patterns that the neuron can send out in response to the appropriate incoming pattern. These firing subcentres are then further supposed to be 'set off' by an appropriately high level of some ionic substance. If the microtubules are capable of rapid transmission of some ion (perhaps calcium itself), a high concentration of the ion could occur at the point at which the microtubules met and formed bridges. Thus, a particular pattern of incoming pulses would initiate ion transmission within the microtubules originating at the relevant dendritic synapses, and the coincidence of the arrival of the ion from a number of these interconnected microtubules at more or less the same time would initiate a particular firing pattern in the neuron.

Learning, in this paradigm, would be the establishment of appropriate networks of microtubules. The known speed of assembly and disassembly of these structures fits our requirements in this area.

However, this simple model does not take into account the fact that the incoming information at each synapse is itself a pattern of pulses, with the information content coded as the time locked frequency of firing. The mere event of a pulse arrival at each synapse cannot be the necessary and sufficient event to initiate the ion transmission within the microtubule network. We require some means of selectively reacting to the micropatterns (time locked frequencies) at each synapse.

In addition to the network of microtubules, it is therefore necessary to have specialized activity and associated molecular structures at the pre- and postsynaptic sites and at the axon hillock.

The next section presents a theoretical statement of the synaptic events associated with neuronal information processing, and the molecular mechanisms which may underlie them.

The Cholinergic Synapse and the Memory Molecule

We have previously noted that information is encoded within the CNS by frequency modulation. The time intervals between nerve pulses vary from moment to moment, and reflect the stimulus strength being presented to the receptor cell (or cells) at the afferent end. Stronger stimuli cause higher frequencies of firing, whilst weaker stimuli are associated with lower frequencies.

The faster axonal natural frequencies tend to be about 250 Hz, corresponding to a 4-ms time period between pulses. Nerve pulses can often be observed to appear in closely spaced groups along a single fibre. These groups of pulses are called pulse trains. Within these pulse trains, the slower frequencies seem to be of the order of 55 Hz, corresponding to an 18-ms time period. In addition to these trains of pulses, individual spikes can be seen from time to time.

Receptor cells can modulate their frequency quite quickly in response to stimulus changes. For example, the sound of the human voice is amplitude modulated by the larynx, superimposing a frequency of about 100 Hz on the higher frequencies. These low frequencies are easily detected (a whisper is a normal voice without the larynx frequencies added) and can be represented by changes in the time intervals between pulses initiated by the cochlear hair cells. The resultant pulse train would thus have a pattern of alternating close and not so closely spaced pulses, representing the 100-Hz larynx frequency, amplitude modulating the higher frequency the particular hair cell is primarily sensitive to.

We believe that a representation of these micropatterns of pulses is capable of being stored within the brain in a permanent form, so that if the same stimulus occurs again, it can be recognized.

To continue the above example, we note that human language is made up by stringing together groups of common sounds, called phonemes. Each phoneme is characterized by the frequencies found within it, which are broadly organized into bands (formants). The continuous sound of speech is in fact made up of these strings of phonemes which change rather abruptly. The average duration of a phoneme is about 100 ms, but ranges from 40 ms or so to about 200 ms. When the cochlea is responding to a particular phoneme, it will be by firing a number of groups of hair cells, corresponding to the formants of the particular phoneme. The firing pattern of each of these cells should in turn reflect the changing amplitudes of each formant over the time course of the phoneme duration.

Suppose, for a moment, that a particular phoneme was to be 'recognized' by a single neuron at some cortical centre. We would have to assume that fibres from the relevant hair cells made synaptic contact with the neuron. However, the likelihood is that our hypothetical neuron makes contact with many cochlear fibres, the majority of which will not be involved in the sound representing a single phoneme. In order to recognize the phoneme, then, the neuron must somehow 'know' that a particular pattern of input fibres is firing, and at the same time, more or less ignore many other cochlear fibres that might be firing at the same time.

There is, moreover, a complication, in that the human voice is never presented in isolation. If we were to listen to a human voice near, say, a waterfall, the 'white noise' of the falling water would be represented as the entire spectrum of audible frequencies, all firing at the same time. How is it, then, that we can pick out the human voice from the background noise of the waterfall?

The answer lies in the pattern of pulses that is generated in normal speaking, primarily by the larynx. These patterns will be quite different from the patterns that would be presented in the same nerve fibres by the waterfall noise. In fact, the random waterfall noise will probably sum at the eardrum with the voice sounds, and in effect it would be somewhat like hearing the voice rather more loudly than usual.

For phoneme recognition to occur, then, the neuron must be selectively able to respond to a particular pattern of incoming pulses along a single fibre, and then be able to tell that a number of such patterns occurred at the same moment in time along a group of input fibres corresponding to the formants in the phoneme. We have already discussed how the neuron might be able to perform the logical 'anding' of the various patterns, by summing an ionic concentration carried by microtubules. How does the recognition of the micropattern occur at the synapse?

To answer this, we need to examine the detailed structure of the synapse, at the molecular level. It is necessary to present a synthesis of both theory and fact, as our current knowledge is still very limited.

Previous publications (AE Hendrickson 1972, Hendrickson and Hendrickson 1980) have presented a detailed model of a possible way in which synaptic recognition might work. The essence of the model is that memory is encoded into short oligonucleotides; specifically a small species of RNA, which we have suggested be called engram RNA, or eRNA. The following sections represent a revised statement of some of the details of this model, incorporating some important recent experimental work which has greatly increased our knowledge of the detailed structure of the synapse[1].

1 We have previously noted (Hendrickson and Hendrickson 1980) that the size of eRNA would be specific to a given species, correlating with certain neurophysiological parameters. Some specific predictions have been made for a number of species (AE Hendrickson 1972), and in 1977 a research group at the University of Bath attempted to find the predicted eRNA molecules in rats and human brain material. It was established that there were in fact well-

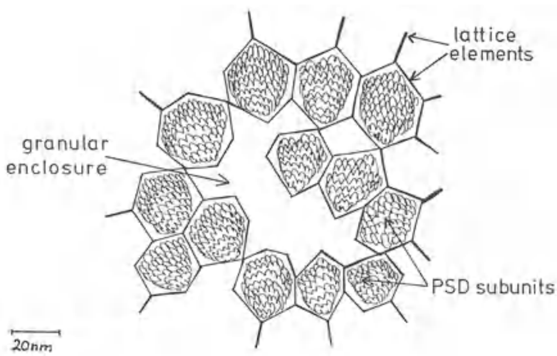


Fig. 2. Schematic drawing of postsynaptic density (PSD), adapted from Matus (1981)

The Postsynaptic Density

The cholinergic synapse can easily be recognized in micrographs by the synaptic vesicles which cluster just inside the presynaptic membrane. The pre- and postsynaptic membranes are separated by a relatively wide gap (about 200 Å) and with appropriate staining techniques, a dark band, called the postsynaptic density can be seen forming a thickening of the membrane on the side opposite to the synaptic vesicles.

A few years ago, Matus and Walters (1975) were able to isolate the postsynaptic densities (PSDs) from brain homogenates by treating synaptosomal plasma membranes with detergents. This development made PSDs directly accessible to biochemical analysis and has provided new insight into the way they are constructed.

Briefly, the PSD appears to be composed of a planar array of small hexagonal-shaped compartments. Each of these subunits falls within a narrow size range, with a median diameter of 180 Å. Groups of these compartments are seen to surround one or more islands of fine granular material set amongst the subunit array.

Figure 2 is a schematic drawing, adapted from Matus (1981), showing the hexagonal arrays surrounding an island of fine granular material.

Westrum and Gray (1977) have shown that there is a close association between the PSDs and microtubules, and this has been confirmed by the identification of tubulin in isolated PDSs by biochemical methods (Kelly and Cotman 1977) and by whole brain immunohistochemistry, using an antiserum as a histochemical stain to reveal the location of the protein (Matus et al. 1975, Walters and Matus 1975).

A number of other substances have been identified at PDSs (Matus 1981). The most interesting of these (for our purposes) are calmodulin, which has been confirmed as an intrinsic PSD component (Grab et al. 1979, Lin et al. 1980, Wood et al. 1980), actin, which has been identified by amino acid composition and isoelectric focusing (Blomberg et al. 1977, Matus and Taff-Jones 1978), and a high molecular weight microtubule associated protein (Matus 1981).

---

defined peaks of RNA of the different predicted sizes in these two species. Having made this discovery, the research became directed towards proving that this RNA was indeed engram RNA and not degradation products of high molecular weight RNA produced during the isolation procedures. A recent experiment (W Whish, personal communication, 1981) has shown that pure radioactive RNA (molecular weight 25000) added to the medium in which the brain is homogenized survives the extraction procedure without significant hydrolysis, thus indicating that the low molecular weight RNA found in vivo is unlikely to have been produced by the degradation of cellular RNA during the isolation procedure. Some further experiments are now being carried out to establish this for human as well as rat material, and to show more directly that the RNA found is related to memory in the way we have hypothesized. These experiments will be reported in detail elsewhere.

We have previously noted that calcium might have a role in the microtubule network linking the synapses with the initial axon segment. There may be other ionic substances involved as well, but it seems unlikely that calcium has no role to play, as it is known to regulate microtubule formation and decomposition in vitro. As will be seen in the next section, calcium may also have an important role to play at the synapse in the recognition of particular pulse patterns. The presence of calmodulin in the PSD clearly implies a functional role for calcium within the PSD, and this gives some support to our model.

Actin, of course, is known to be one of the two main muscle proteins, and at first sight, it might seem strange to find this particular protein at this site. However, it turns out that there is a definite theoretical need for actin, or something like it, at the postsynaptic site.

Our most recent statement of our theory (Hendrickson and Hendrickson 1980) hypothesized that eRNA was to be found under the postsynaptic membrane, attached to a microtubule end, which was then thought to be able to rotate during the arrival of a pulse train at the synapse (for reasons detailed below). The existence of actin, however, provides a more elegant way for our eRNA to travel, and of course, has at least the confirmation that actin is known to be present and have the function of kinesis.

There is one other interesting aspect about the PSD that has been noted by Matus and Taff-Jones (1978). They found that the interiors of the hexagonal PSDs seemed to be highly hydrophobic, which they concluded because the aqueous heavy metal salts of a negative stain that was used were unable to penetrate the hollow interior of the polygons. This we find significant because we believe these PSD units to be containment cells for the eRNA. In order for the eRNA to work in the way we suppose it to, it is desirable (and perhaps necessary) that there be a minimum of water present.

The Memory Molecule

Within the PSD, we suppose the memory molecule to reside. This is thought to be a small molecular strand of ribonucleic acid (RNA). Engram RNA is thought to be rather different from most other RNA within the cell in at least two respects.

The first way in which eRNA is different is in size. At the moment, the smallest known species of RNA have molecular weights of about 25000, corresponding to some 75–80 of the nucleotide subunits of RNA. In contrast, eRNA is thought to vary between 21 nucleotides in size ($\sim$7000 daltons) down to, perhaps, 9 or 10 nucleotides ($\sim$3000 daltons), the actual number being dependent on the species of animal.

RNA is a macromolecule, made up of connected subunits, which are collectively called nucleotides. There are four such nucleotides which can be found in RNA, namely guanine (G), cytocine (C), adenine (A), and uracil (U). Each of these subunits is a planar molecule, consisting of one or two ring structures. The planar base is attached to a sugar-phosphate backbone, which holds the structure together.

As a three-dimensional structure, RNA can be viewed as something like a stack of plates. The distance along the sugar-phosphate backbone between each nucleotide base is approximately 7 Å. The backbone of the molecule has a negative charge, and the flat plates of the individual bases have highly hydrophobic faces, which allow them to stack together without water molecules intruding.

Around the edge of the bases are points at which the molecules can form hydrogen (H) bonds. In the sister molecule to RNA, known as DNA, these sites are used to hold two complementary strands of DNA together. This is, of course, of fundamental importance to all known forms of life, as DNA is the molecule which carries our genetic blueprints. The H-bond sites which hold the double-stranded DNA together are the means by which the DNA molecule is

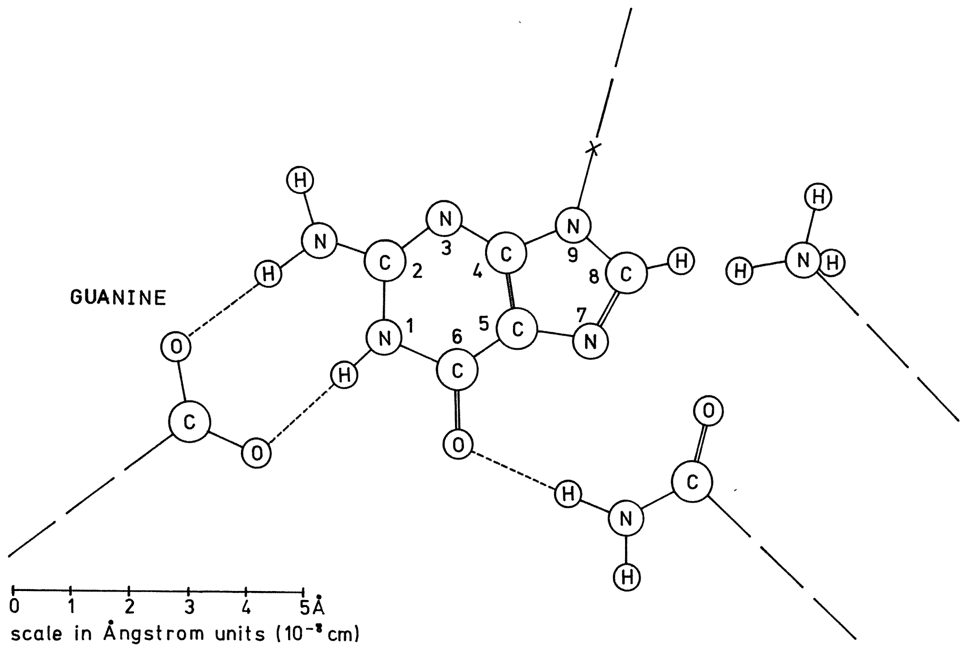


Fig. 3. Sectional view of guanine attached to a hypothetical substrate. Three relatively strong H-bonds are formed, giving this nucleotide the strongest attachment to the binding site

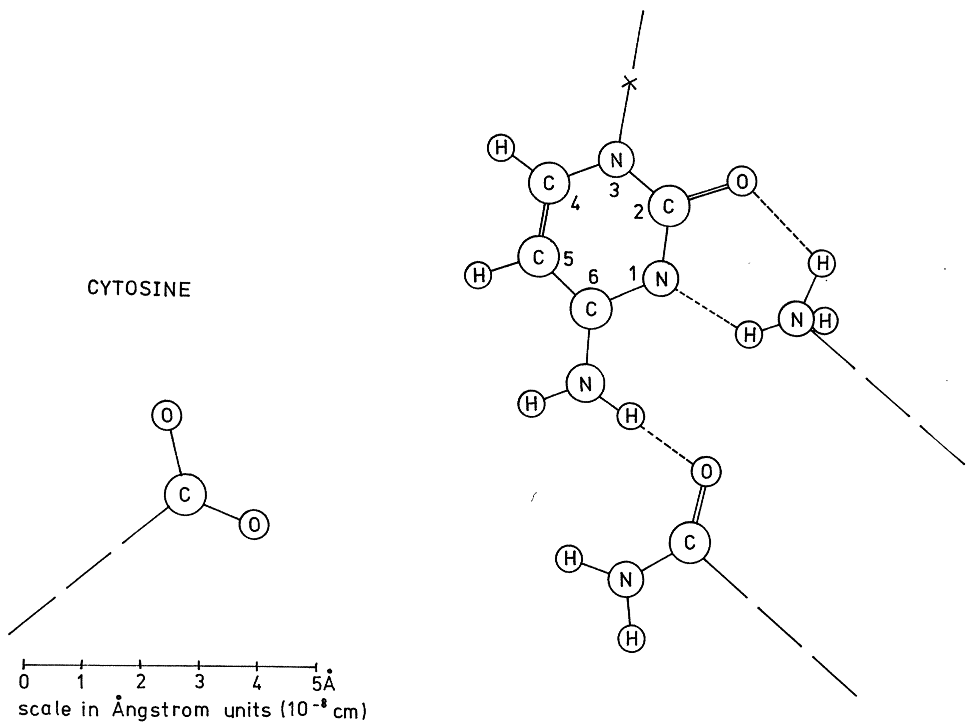


Fig. 4. Sectional view of cytosine attached to a hypothetical substrate. One strong and two weaker H-bonds are formed to the amino acids in the substrate grove, giving this nucleotide the second strongest attachment

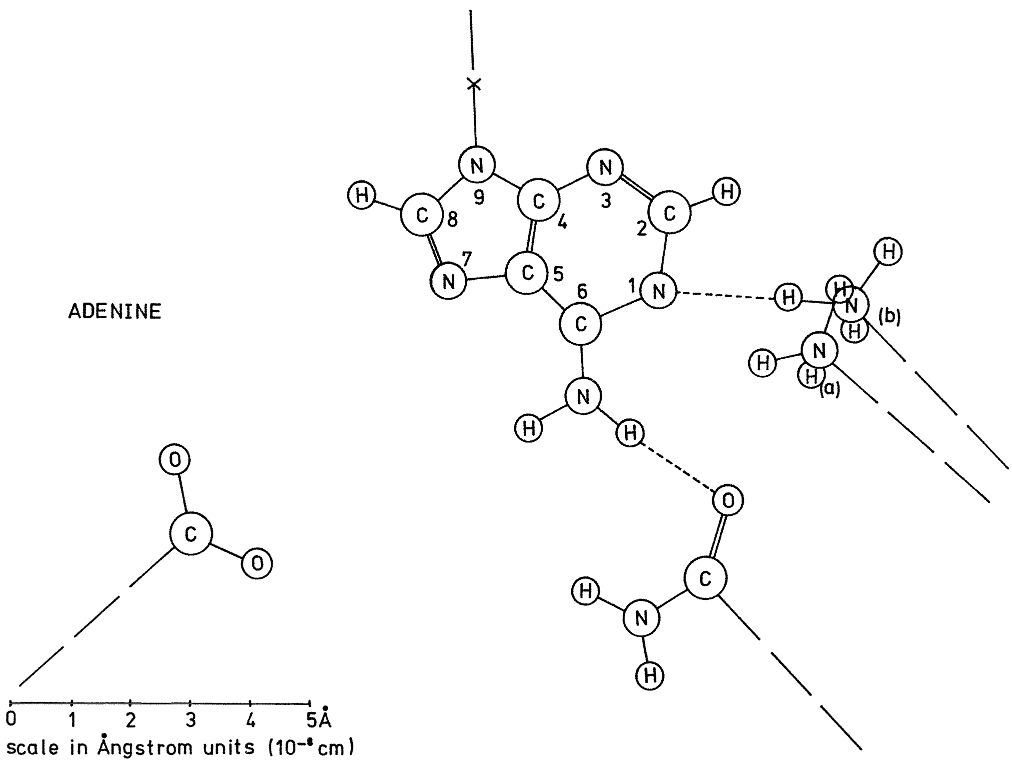


Fig. 5. Sectional view of adenine attached to a hypothetical substrate. Two strong H-bonds are formed, giving this nucleotide the third strongest attachment to the universal binding site

able to copy itself, which is at the core of our life processes.

We suggest that these same H-bond sites have a role to play in RNA as well as DNA, and form the basis on which RNA is able to act as a memory 'template' at the synapse.

Our model pictures eRNA sitting within the PSD. In the human, the eRNA strand is thought to be some 21 bases long, corresponding to a molecular weight of 6700. It would be about 150 Å in length, and therefore fit comfortably in the 180 Å diameter interior of the PSD. [Our 21-base hypothesis was made over 10 years ago (AE Hendrickson 1972)]. The molecule would be attached to a special substrate 'binding site' which would have the special property that would allow it to form H-bond attachments to the bases along the length of the eRNA. The H-bonding attachment would occur no matter what the sequence of bases was, but the number and angles of H-bonds that were formed would vary as a function of the identity of each base.

We picture G as forming three H-bonds to the substrate molecule. Each of these three H-bonds would be straight. C would also have three H-bonds, but two of these would be bent, which has the effect of weakening them. Next, we have A, with two straight H-bonds, and finally U, with at least one of its two H-bonds at an angle.

The importance of the model, as described above, is that it allows the eRNA molecule to have any arbitrary sequence of bases along its length, and secondly, that the bases are not attached to the substrate with the same degree of strength.

Figures 3–6 present one suggested way that the four nucleotide bases might form H-bonds to some universal binding site. These scale drawings represent the end results of some amateur attempts we have

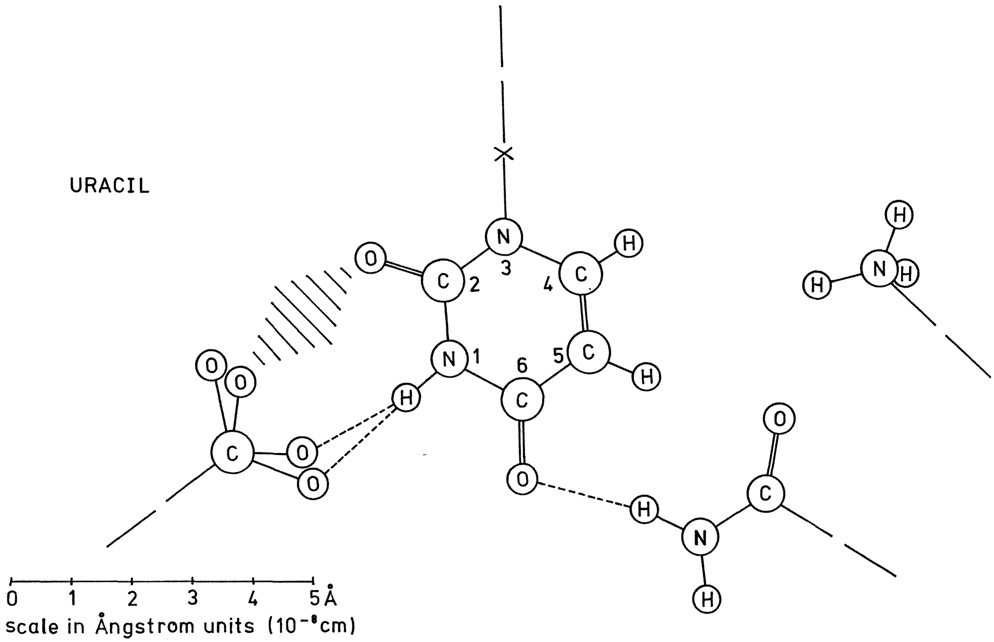


Fig. 6. Sectional view of uracil attached to a hypothetical substrate. The combined strength of the two bonds must be weaker than the two bonds of adenine. Here we show a rotation of the COO-group to effect a bond angle out of the plane represented by the drawing

made to study the stereochemical constraints on our model, and are intended only to be suggestive.

If we imagine that we could somehow grab one end of an eRNA strand and tear it loose from its substrate, we would feel it come away in an uneven manner, depending on the base sequences along its length.

We now try to picture this structure, and what happens when a train of nerve pulses arrives. Just below the postsynaptic membrane, which covers the PSD cell, sits the backbone of the eRNA molecule. Initially, however, it is not completely exposed. The interior of the PSD is thought to be divided into two subchambers, with a 'shield' like structure, which prevents the entire backbone of the eRNA molecule from being exposed to incoming sodium at the same time.

The eRNA is attached to a substrate molecule, thought to be a microtubule-associated protein (MAP). The MAP itself is pictured as being closely associated with a short strand of F-actin. The actin is initially in the 'contracted' state. In muscle, calcium is known to bring about the contractile state, and we therefore suppose our 'initial' state to have a high concentration of calcium in one of the two subchambers. Figure 7 presents a schematic drawing of these structures.

Now a pulse arrives at the synapse. The synaptic vesicles at the presynaptic side spill out their contents of the 'transmitter' molecule, acetylcholine (ACh). The ACh has the known function of opening up special channels connecting the PSD to the synaptic cleft. These channels allow sodium to enter the postsynaptic side, which it does because of a 10 to 1 concentration gradient between the two sides of the membrane.

The sodium ion has a positive charge, and we therefore imagine that there is an interaction between the positive charge of the sodium and the negative charge along the eRNA backbone. The effect of this is that the sodium weakens the H-bond attachment of the eRNA to its substrate, by pulling it

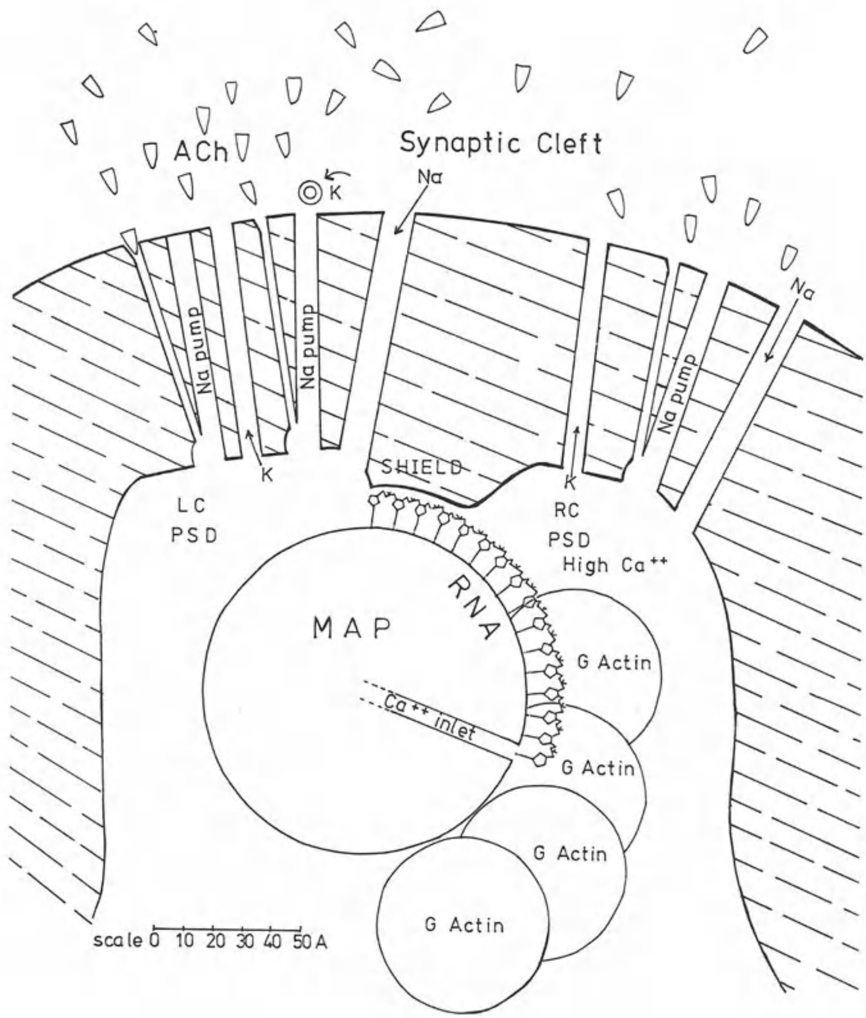


Fig. 7. Hypothetical sectional view of the interior of a hexagonal PSD chamber. Na and K channels and the Na/K pump are shown, allowing the temporary build-up of Na above the eRNA, which is attached to a rotating globular microtubule-associated protein. The motility is provided by actin, with Na antagonizing the effect of Ca^{++} in the right half of the PSD chamber

away. As H-bonds are stretched, they are known to weaken, and it follows that if the positive pull of the sodium is strong enough, it might actually detach one or more of the bases of the eRNA.

However, the entire length of the eRNA backbone is not exposed to the increased sodium concentration. The most exposed of the bases might or might not be detached. On the supposition that it is detached, or at least greatly weakened, we then picture the eRNA molecule as moving a bit, coming out from under its shield by virtue of some relaxation on the part of its actin carrier. In muscle, calcium is responsible for the initiation of contraction, and sodium is a competitive inhibitor, and antagonizes the calcium effect (Katz 1966). We have merely to suppose that the same holds true in the PSD, and that the influx of sodium antagonizes the effect of the calcium, weakening the contraction of the actin. The net effect, then, would be that another length of eRNA backbone would come out of its

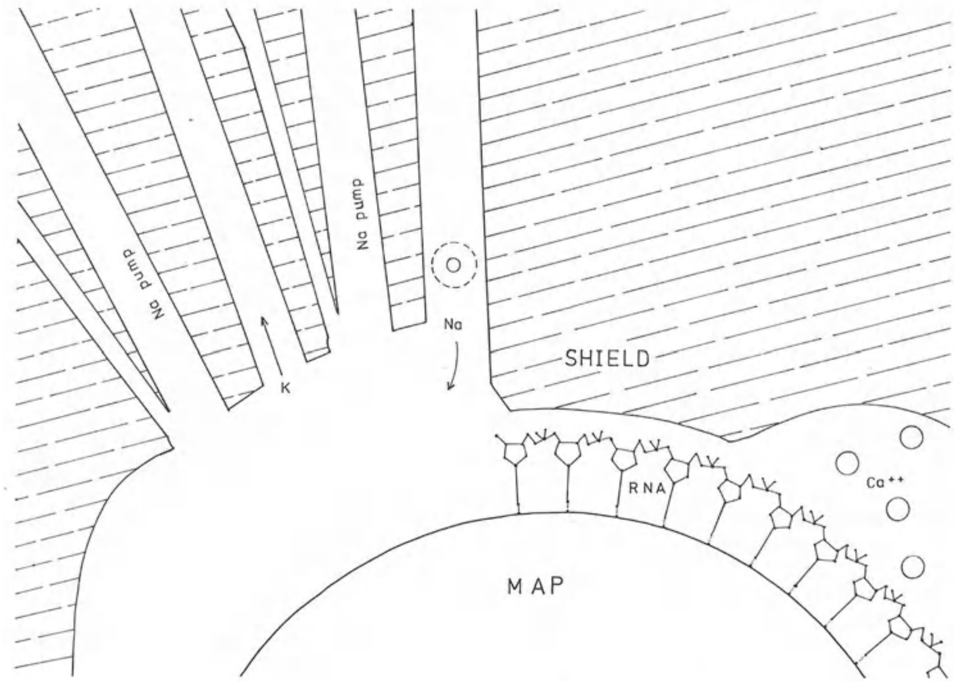


Fig. 8. Detailed view of the eRNA attached to the MAP within the PSD. This shows the position of the molecules just prior to the arrival of a pulse train at the synapse

shelter, and be in place for the arrival of the next nerve pulse.

At this point, we must be reminded of the fact that the arrival of the next nerve pulse at the synapse is not a scheduled event. It is dependent on the stimulus strength, and the train of pulses will arrive at various time intervals, forming a sort of pattern, which is stimulus driven.

The sodium which has been allowed to enter the PSD chamber does not, however, remain in the newly heightened concentration, as a molecular 'sodium pump' begins at once to pump it back into the synaptic cleft. Thus, while we are waiting for the next pulse to arrive, the sodium concentration is dropping inside the PSD chamber.

What follows now is a battle of sorts. The sodium may, on the one hand, counteract the calcium, and draw the eRNA out of its shelter, and at the same time work on pulling the strand off its substrate. Whether it is able to succeed or not depends on two things. Firstly, the sequence of nucleotide bases along the eRNA strand will partly determine the outcome, as the attachment strength will vary from base to base because of the differences in H-bond numbers and angles. Secondly, the time intervals between the pulse arrivals is important, as we picture the sodium pumps working at more or less a constant speed. The actual sodium concentration at any point in time is therefore a function of the time which has elapsed since the arrival of the preceding nerve pulse. A train of closely spaced pulses will cause a high level of sodium to build up, whilst a series of widely spaced pulses will result in a slow build-up of sodium within the chamber.

The sodium concentration will thus rise and fall throughout the course of the pulse train, and if the changing concentration 'matches' the eRNA nucleotide sequence such that the more strongly attached bases encounter the highest concentration levels, the eRNA strand will be completely torn away from its substrate. The figures below

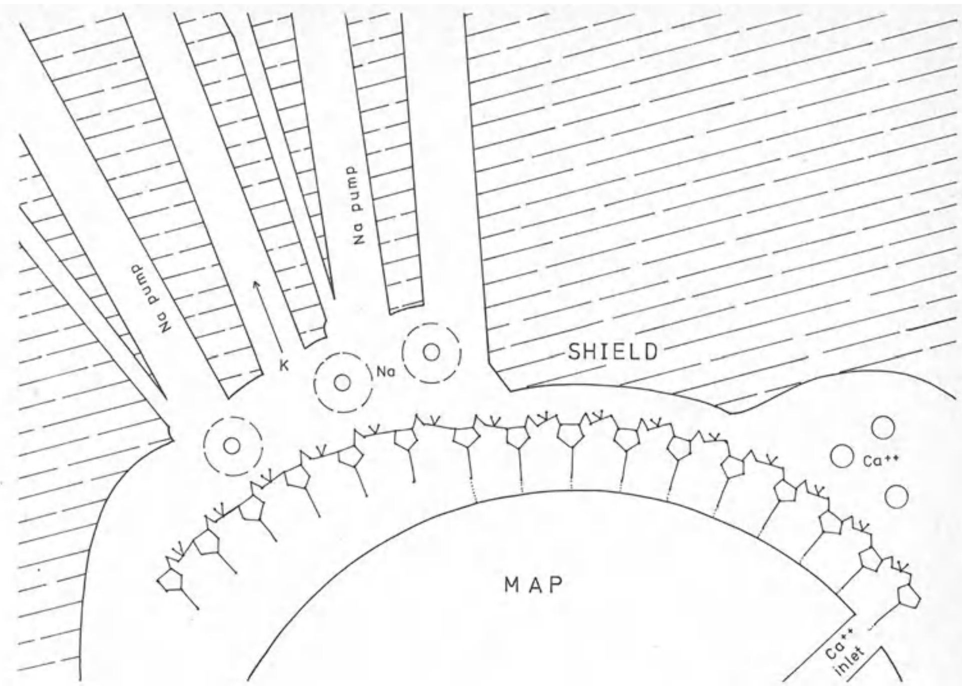


Fig. 9. View of the interior of the PSD after seven pulses have arrived and successfully broken the H-bonds of six of the nucleotides. The bonds of the seventh are extended and very weak. The MAP has rotated, and the Ca^{++} inlet can be seen to the right

show the hypothetical interior of the PSD, with the eRNA attached to an MAP/actin complex, interacting with incoming sodium ions during the arrival of a pulse.

We picture this dynamic interaction between pulse train time intervals and eRNA base sequences in such a way that it is thought that the eRNA effectively codes, or acts as a template, for a subset of possible pulse train patterns. (We have previously further hypothesized a one-to-one correspondence between time intervals and bases, supposing there to be only four possible time intervals. The evidence supporting this, however, is not clear, and in any event the refinement is arguably elegant but unnecessary).

A template 'matching' between the eRNA and the pulse train intervals is thought to have the end result that the eRNA is completely detached from its substrate. What happens then is that the ionic 'transmitter' (whatever it is) enters a microtubule ending that has become exposed because of the eRNA detachment. The synapse has effectively 'recognized' an incoming pulse train. We believe that this molecular event is, in effect, the basis for the retrieval of memory.

If several sites on the same neuron also 'recognize' incoming pulse trains, and the microtubules are ones that are associated by means of the cross-bridges that have previously been described, a build-up of the internal ionic transmitter will take place in the initial axon segment, and the neuron will fire some pattern of pulses of its own. We suppose each neuron to have a repertoire of pulse patterns that it can send out, which become associated with particular input pulse train combinations through learning.

The fact that water molecules would interfere with the three-dimensional stacking of RNA has been noted, and it is probably the case that the intrusion of water between the H-bond sites at the base of the eRNA, and the hypothetical substrate, would prevent proper binding. If this is the case, we

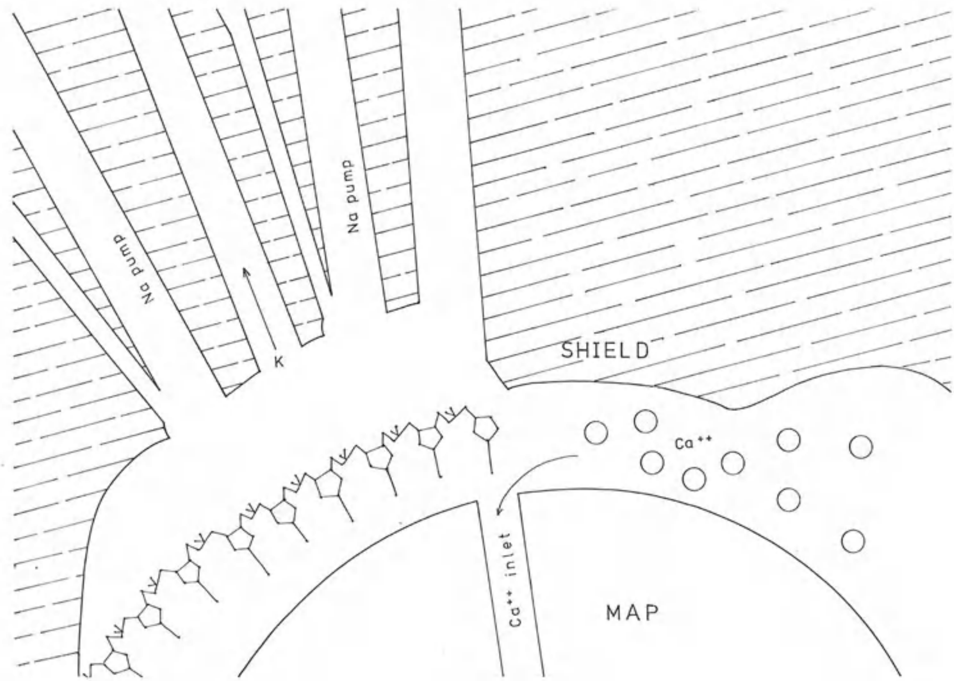


Fig. 10. The pulse train has stripped off the eRNA completely from the MAP substrate, exposing the Ca^{++} inlet channel, which is now allowing Ca^{++} to enter. This constitutes recognition of one pulse train pattern. If a number of such patterns are recognized at the same time at other synapses on the same neuron, the combined build-up of Ca^{++} at some site in the axon hillock will initiate a particular firing sequence

would expect the PSD site to be hydrophobic, and it is therefore interesting to note again the finding of Matus and Taff-Jones (1978) that the PSD interiors are probably highly hydrophobic.

The Role of Acetylcholine at the Synapse

We have briefly noted the fact that ACh is released into the synaptic cleft from large (400–500 Å diameter) vesicles, where it momentarily acts to open the sodium inlet channels.

However, many more ACh molecules are released into the cleft than are needed for the purposes of opening these sodium channels. The ACh molecules are known to bind momentarily to the postsynaptic membrane in some way, and they are then acted upon by a very fast-acting enzyme called acetylcholinesterase (AChE). The action of AChE is to split ACh itself into the two parts, choline and acetate.

We now pose some simple questions. Why does this happen, and what function does it serve? If ACh is a 'transmitter' molecule, what indeed is it transmitting?

We have suggested that an additional role (perhaps the 'main' function of the ACh molecule) is as a regulator for the sodium pump.

The sodium pump itself is a fast-acting mechanism. It is not yet known what the exact molecular structure of the pump is, but many facts about its performance are known. The speed of the pump is such that, along the axon, it is able to re-establish the original sodium/potassium ratios within 4 ms after the passage of a nerve pulse. Indeed, most of this is accomplished within a single millisecond.

We suggest that the same molecular pump mechanism that is to be found along

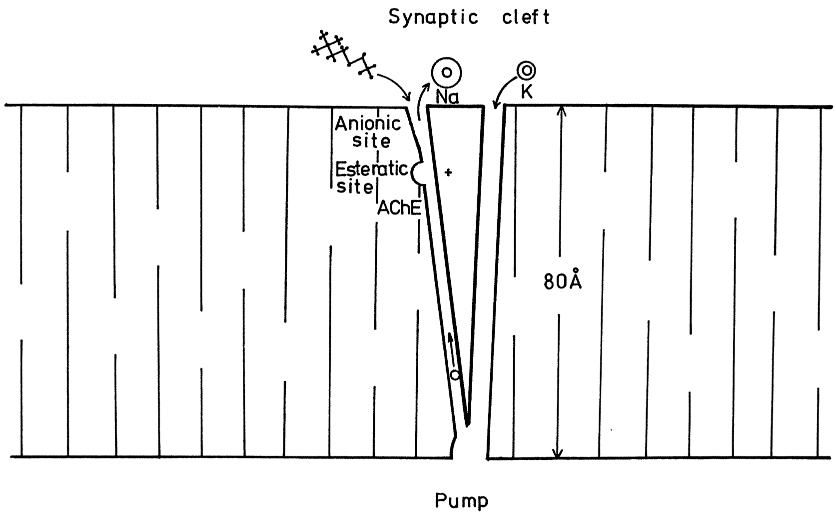


Fig. 11. Schematic sectional view of the postsynaptic membrane, showing an ACh molecule just about to enter the sodium egrees channel of the sodium pump. The highly polarized ACh molecule is guided into place by a positively charged site at the entrance to the channel

the axon length is used at the synapse, but that it is somewhat inappropriate for the job because it is too fast. However, the ACh molecule slows it down by momentarily blocking the sodium egress channel. We picture a single ACh molecule blocking each channel, and then rapidly being chopped by the action of AChE. Usually, as soon as the choline is expelled from the mouth of the channel, its place is taken by another ACh molecule. On occasion, however, a sodium ion is allowed to slip out before another ACh molecule can get into the channel.

The net effect of this is to slow the sodium pump action down and also to provide a smoother, more linear recovery of the original internal sodium concentration. This probably has important consequences in the template action of the pulses and the eRNA molecule.

Figures 11–13 show a schema of the sequence of the arrival of the ACh molecule, its momentary blocking of the sodium pump, and its subsequent hydrolysation.

The evidence supporting our picture of this putative action of ACh is derived from a number of considerations. Firstly, there is the shape of the molecule, which is somewhat like a wedge. It is about 9 Å long, and the thick end of the wedge is about 4 Å wide. The narrow (front) end easily slips into the sodium channel, and the thick back end cannot enter. The molecule has a convenient polarized charge, with a negative charge at the front and a positive charge at the back. Thus, the charges act as a guidance system, with the bonus that the positive charge at back expels the choline once the AChE action occurs. The second piece of evidence supporting our suggested role for ACh is the well-documented effects of certain agents which antagonize the actions of AChE. When such a substance (such as eserine) is administered at the neuromuscular junction, the muscle contraction is prolonged (Katz 1966). Evidently, if the AChE cannot act, the sodium cannot be pumped out, and the muscle remains in a contractile state. This clearly implies that the ACh is blocking the egress of the sodium. We have now simply to suppose that what is true at the neuromuscular junction is also true within the CNS. (We think that the primary ACh function is the same at the neuromuscular junction and serves to provide a smooth recovery of muscle cells after contraction.)

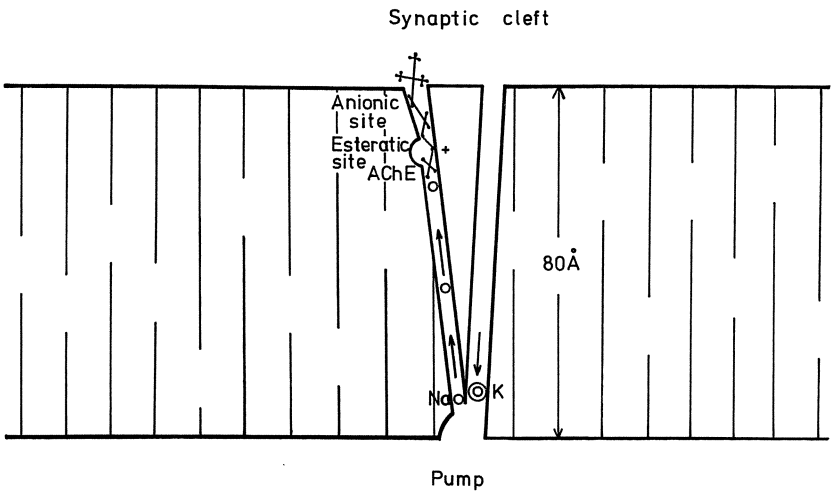


Fig. 12. View of the postsynaptic membrane, with an ACh molecule in place, blocking the action of the sodium pump

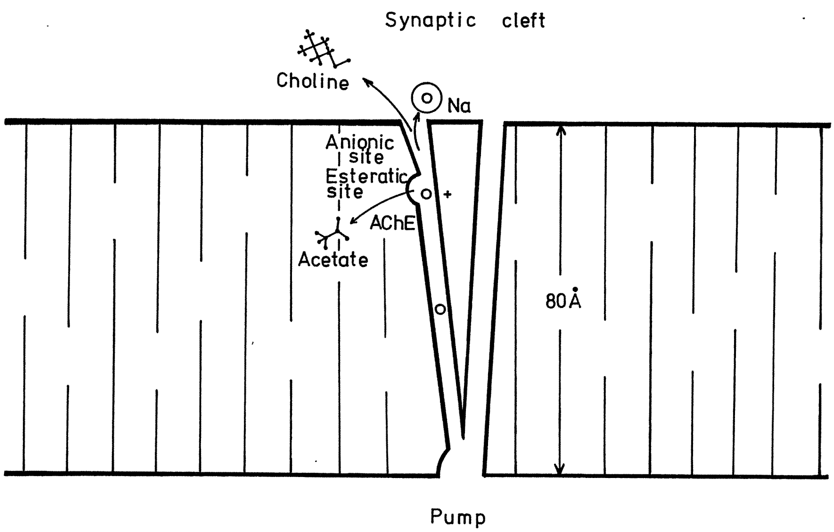


Fig. 13. View of the postsynaptic membrane after the action of AChE. The ACh molecule is now hydrolysed, and a sodium ion has been able to escape. The site is now available to another ACh molecule. The overall effect is to slow the egress of sodium from the PSD chamber

Learning and the Formation of Memories

eRNA is unusual in that it is much smaller than other species of RNA that have been studied so far. It is also quite unusual in another respect, and that is in the way it is formed.

Some eRNA, probably a minority of the total in the adult animal, is formed in the conventional way, by transcription from our DNA.

This 'preformed' RNA allows animals to be able to recognize and respond to stimuli the very first time they are encountered. This is the molecular basis of what we normally call 'instinct'. Instinct would comprise not only entire sequences of behaviour, but would also include the important attribute of our innate ability to recognize reinforcing stimuli.

However, much of our adult behaviour is the result of learning, which we see as the formation of the paired association between a recognized input, and some behavioural output. We have already mentioned part of this learning process as being the formation of microtubule cross-bridges in the initial axon segment, where the initiation of firing sequences occurs.

The other molecular aspect of learning is the formation of eRNA sequences that will match stimulus input pulse patterns when they re-occur. How is this accomplished?

We believe that eRNA can be formed by the action of a special 'learning enzyme' acting on the appropriate precursor RNA subunits. No DNA transcription is involved in this formation at any time.

The hypothetical 'learning enzyme' may possibly be a known substance called polynucleotide phosphorylase (PNP). PNP may not be the actual learning enzyme, but it will suffice for the purposes of our description if we assume that it is. (As far as we know, PNP has not yet been identified in eukaryotic cells.)

We must now imagine another molecular structure something like the one we pictured in the PSD. Again, we have the universal binding site, which can form H-bonds to any of the four nucleotides it encounters. However, instead of just one of these binding sites, we have an array of four of them, sitting side by side. Like our previous model, we picture this array sitting on actin.

Attached to this array will be the precursors, which are the nucleotide diphosphates. We have selected nucleotide diphosphates because they are known to be the precursors of RNA that can be formed by PNP. The diphosphates are very much like the nucleotide bases in the RNA molecule itself, except that they are not attached, and they have an extra phosphate group. The negative charge associated with each nucleotide diphosphate is thus double that of the base in the context of RNA.

We picture our array of four substrates as containing a series of identical rows of the diphosphates, in the order G, C, A, and U. The array itself is found at the synapse, and we suggest that it could be found in the islands of 'fine granular material' that are surrounded by the hexagonal PSD cells, described by Matus (1981).

We believe that sodium is released into this area by the arrival of incoming pulse trains, and that, in most respects, the molecular actions are very similar to our previous description of the recognition model. However, closely spaced pulses will detach complete rows of the diphosphate nucleotides, while the longer pulse intervals will only be sufficient to free the weakly attached uracil diphosphate. Intermediate pulse intervals will detach the diphosphates of U and A, or U, A, and C.

Let us suppose that a pulse train has just arrived, and corresponding to the pulse intervals, a number of diphosphate nucleotides have been released in each of the rows. They are now more or less floating in the medium just above the point at which they were released. Now let us suppose that a quarter of a second later, the animal receives a reinforcing stimulus of some kind. He recognizes it in the same way he recognizes any other stimulus, except that the eRNA used was transcribed from DNA. However, the action taken when the reinforcing stimulus arrives is to cause a local release of PNP. We further suppose that the PNP is released adjacent to the free-floating diphosphates, and as we have shown in Fig. 15, somewhat to the left of them. The PNP now goes to work, and goes along the free diphosphate nucleotides, taking the leftmost when there is more than one, and knits them into a string. It is somewhat difficult to imagine this event, but we know that it does happen in vitro.

At the end of its knitting action, the PNP will have formed a strand of eRNA that corresponds in the way we have described to the input stimulus pulse pattern. The

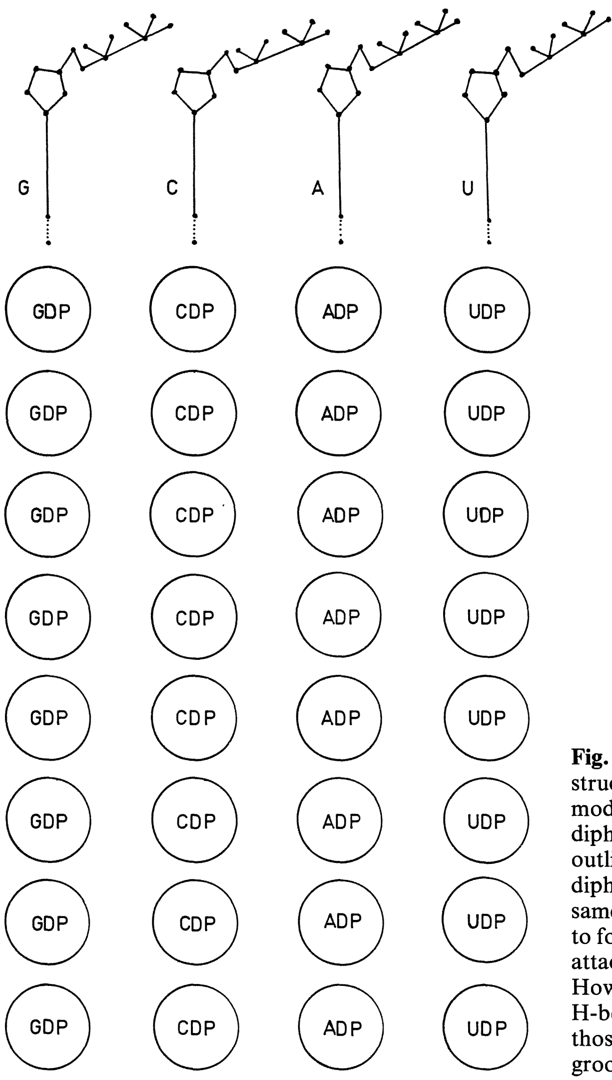


Fig. 14. Schematic diagram of the molecular structure hypothesized in the encoding model. The top of the figure shows the diphosphate form of the nucleotides in outline. There are four columns of diphosphates, each column being of the same type. They are thought to be attached to four substrate grooves, each capable of attaching only one sort of diphosphate. However, the relative strengths of the H-bond sets formed would be the same as those formed in the universal substrate groove utilized by the eRNA strands

eRNA will be capable of 'recognizing' the same pulse pattern that fathered it when and if it re-occurs at the same synapse.[2] After the eRNA has been formed, we picture it migrating to one of the adjacent PSD hex cells, 'searching around' until it finds one with no eRNA already in place.

When the new strand of eRNA has found a home, it begins to perform its duty of scanning incoming pulse trains. When recognition occurs, the ionic transmitter is released into the microtubule that is attached to the hex chamber. At the distal end of the microtubule, we imagine the arrival of

2 The eRNA strand would actually be capable of responding to a class of stimuli, rather than just the progenitor pulse train. This is because the shortest pulse intervals (resulting in high sodium concentrations) will cause any of the nucleotide bases to be detached. For instance, one could have a 'universal' pulse train, consisting of a sequence of very closely spaced pulses, that would be able to strip off any eRNA strand. Such pulse trains would be caused by very strong stimuli. Thus, for example, an unexpected firecracker going off nearby would cause such pulse trains, which would have the effect of causing many association circuits to fire at the same time. We experience this as the 'startle reflex'.

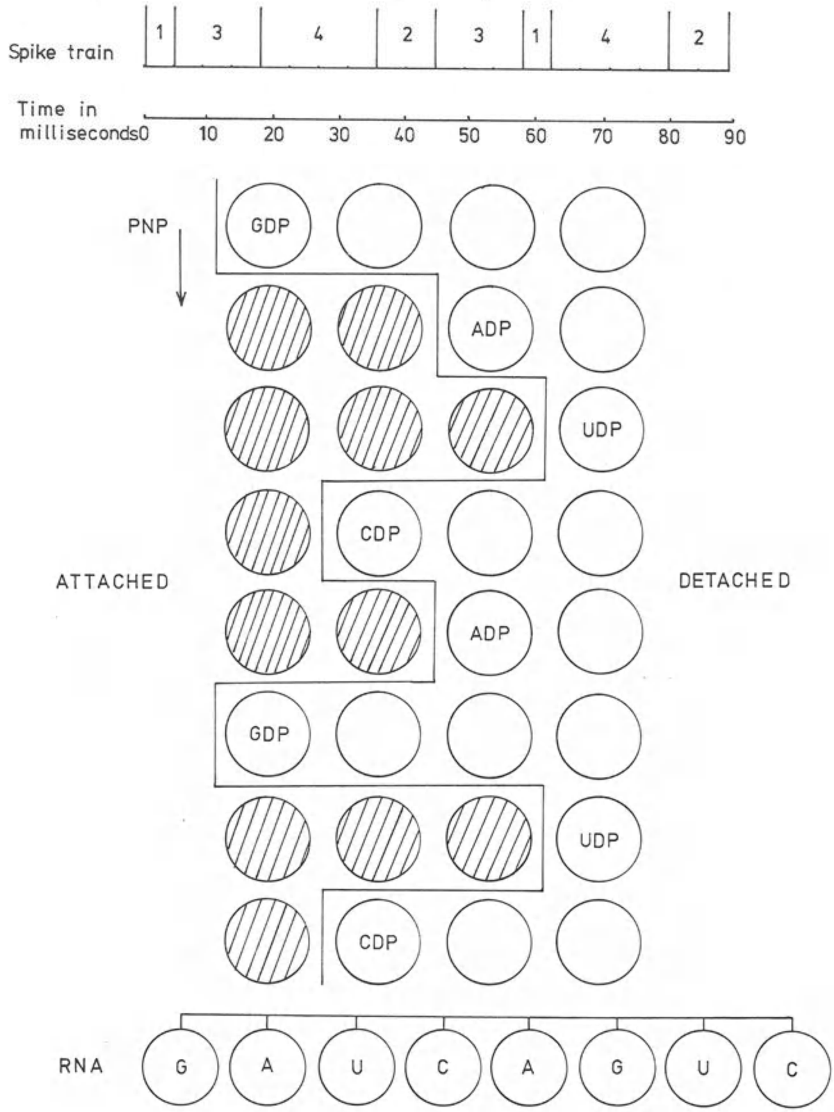


Fig. 15. Companion figure to Fig. 14. The top of the figure shows an oscillograph of a nine-pulse pulse train, with one of four possible time intervals between each pulse. A time scale is shown below the pulse train; in this example, the pulse train is 90 ms long. Below the time scale is the matrix of diphosphates, with eight rows and four columns. Each interval of the pulse train affects one row of the matrix, detaching from one to four diphosphates. The diphosphates remaining attached are shaded. Immediately after the passage of the pulse train, PNP is conditionally released at an adjacent site. If this occurs, the PNP will act on the leftmost of the freed diphosphates, concatenating them to form a new strand of eRNA. A schematic representation of the new eRNA strand encoding the pulse train at the top of the figure appears at the bottom of the figure

the ionic substance as predisposing the microtubule to form cross-bridges and associations with other microtubules that are discharging ions at the same time. In this way, a network capable of recognizing a complex stimulus within a single neuron has been formed.

This, we believe, is the basis of learning.

Pulse Train Recognition, Errors, and Reliability

The detailed molecular models we have presented so far have shown how we might be able to form memories of events that were originally neutral, but which come to have meaning because of their pairing with reinforcing stimuli.

If we extend these notions a little bit, and imagine that when a recognition sequence has been formed within a neuron, it is itself able to cause the release of PNP in an adjacent neuron, we have a means whereby a chain of logically related pulse trains can be built up. We see this ability of the neuron to itself release PNP in response to what was originally a neutral stimulus as the basis for all of our complex behaviour. The secondary release of PNP, of course, is well known to psychologists as 'secondary reinforcement', which can easily be demonstrated in laboratory conditions.

As a neuron will itself send out a pulse train in response to incoming stimuli, it follows that much of the internal 'processing' that takes place in the brain is the recognition of these secondary pulse trains, which are one or more steps removed from the primary stimulus. The setting off of one of these pulse chains (as opposed to trains) is a process that we might call by various names – for example, thinking, or in some contexts, problem solving, planning, letter writing, etc. All of these outwardly appearing different sorts of behaviour should have the same internal molecular mechanisms.

We can represent these chains of pulse trains by the simple schema:

$$A \rightarrow B \rightarrow C \rightarrow D \cdots \rightarrow N \rightarrow$$

Each of the letters in the diagram above represents a pulse train, which is thought to be recognized, thereby setting off the next pulse train in sequence. These sequences are our associative memories, and in the appropriate context, they represent the behavioural sequences alluded to above.

In the human, each of these pulse trains is thought to be a constant 22 pulses (21 intervals) long. The average pulse interval is of the order of 11 ms, and the average pulse train therefore takes about 230 ms to pass a given point. We can use this average time to work out the likely time taken for the longer sequences of pulse train chains.

Ultimately, each logical chain has an end point, and the number of steps required to reach the end will vary as a function of task complexity. We equate these chains of associated pulse trains as being equivalent to the 'programs' that were discussed at the beginning of this chapter. The pulse train is therefore the 'action subunit' that was mentioned as the basis of the programs.

There is no logical reason why these logic chains would be found to reside only in neural structures with adjacent neurons. Once a pulse was initiated at one point in the chain, the next responding neuron could be some distance away. We would also expect that long sequences of logical processing might well involve 'loops' over the same fibres, and it could well be the case that idividual neurons were responsible for more than one of the chain links. Put another way, a neuron could find itself firing in indirect response to previous firing that it had itself undertaken some seconds previously.

In most cases of manifest behaviour, we believe that it will be necessary for a degree of replication of the logical chains to occur. In some cases, this replication may simply be necessary because the outward behaviour involves large muscle groups, and the final neurons involved in the behaviour might be some number of spinal motor units. The acquisition of what we would term a 'skill' would be the gradual build-up of these replications, until sufficient numbers of them were there so that a particular action was performed smoothly and quickly.

What has been presented so far is a picture of molecular events as they should ideally function. It will be recalled that we have discussed reliability in some detail in fairly abstract terms. We now pose the question, 'What happens when the recognition sequence does not work as it is supposed to work?'

Individual Differences in Molecular 'Errors'

We have discussed in some detail the interaction of sodium with eRNA at the synapse, and noted that the ACh molecule acts to slow down the pumping of sodium back into the synaptic cleft. This has the effect of making the recovery of the original sodium concentration a smooth and relatively linear event, as opposed to the quick exponential recovery that is seen in axonal transmission.

We now refine the above question into the more detailed questions, 'What happens if the sodium is pumped out too quickly?', and conversly, 'What happens if the sodium is pumped out too slowly?' Either of the above two events is an 'error' of a sort. However, they lead to quite different sorts of consequences.

Let us first consider what happens when sodium is pumped out too quickly. In this case, the H-bonds of the nucleotide affected by the most recent pulse will recover more quickly, and in some cases, the H-bonds will not break when logically they should have broken! This means that the nucleotide will remain bound to the substrate molecule, and it has the ultimate effect that the eRNA template cannot match the pulse train time pattern. In turn, the failure of the eRNA template to be removed from the substrate means that the ionic substance will not be released into the microtubule network, and this, of course, means that the neuron will probably fail to fire and send out its associated pulse train. The logical chain, whatever it might have represented in more molar terms, thus comes to a halt. We will refer to this as 'recognition failure'.

Recognition failure can probably occur for many different reasons. The description given above implies a failure with the molecular mechanism at the synapse itself. However, if there were problems with pulse transmission along the axon, we would still expect to have recognition failure at the synapse. The axonal failures could range from the complete dropout (loss) of a pulse, to induction of a pulse that should not have been there. In between these two extremes, we can have pulse speed variations.

Pulse speed variations are only troublesome if the variations affect individual pulses in a train The absolute speed of pulses varies considerably, and is a function of axon size. If all of the pulses in a train were to be speeded up or slowed down by the same amount, it should not make any difference to the probability of correct recognition at the synapse. If, however, one pulse is speeded up, it will increase the time interval between itself and the pulse following it, thus making recognition less likely. On the other hand, if the same pulse is slowed down, it increases the time interval between itself and the pulse preceeding it, and again makes recognition less likely. The fact that in both cases another interval gets shorter does not compensate for the longer interval, because of the way our model of the scanning process works. Thus, we conclude that any variation in the temporal integrity of the pulse train as it passes along the axon is likely to make the failure of pulse train recognition at the synapse more likely.

It is possible that the converse sort of error, an increased probability of recognition, can occur. This would happen if there was a 'slow pumping' of sodium at the synapse itself. If for any reason, the concentration of sodium inside the PSD chamber is higher than it should be throughout the arrival of a pulse train at the synapse, it will increase the probability that the eRNA will be stripped off, indicating a recognition event. Obviously, if the sodium concentration is high enough, it could cause the recognition of a pulse train sequence that should not have caused the eRNA to have been removed. This is 'misrecognition' or 'irrelevant recognition'.

The actual molecular 'failure' at the synapse that might bring about misrecognition could be any one of a large number of possibilities, and it is pointless to extend our model to discuss these possibilities at this time.

The discussion of recognition errors has been pursued because we think it to be the most important aspect of the theory and model presented here as it pertains to intelligence. We believe that all individuals can be characterized by the extent to which recognition failures occur or do not occur. This characterization can be expressed as a single parameter, a probability of recognition occuring when it should occur.

The actual parameter as it might apply at a single synapse need not apply to other synapses, but what seems to be likely is that the distribution of individual synaptic probabilities will have a mean, characteristic value for the individual as a whole.

The greater or lesser characteristic probability of correct pulse train recognition at the synapse for given individuals we believe to be the biological basis of intelligence.

Recognition Probability and Intelligence

The preceding section has presented the paradigm of a logically related chain of pulse trains, forming a chain:

$$\mathrm{A} \rightarrow \mathrm{B} \rightarrow \mathrm{C} \rightarrow \mathrm{D} \cdots \rightarrow \mathrm{N} \rightarrow$$

with the final arrow indicating that the final pulse train must be recognized as well for the chain to have some ultimate purpose.

Let us call the characteristic probability of correct recognition for an individual 'R'. The converse probability of recognition failure is therefore $1-R$. R is to be thought of as the probability that just one of the synaptic recognitions will succeed, as opposed to the probability that the entire chain will be recognized correctly.

If we assume that each synapse has the same value of R, and that the probabilities are independent, we specify the probability that a chain of 'N' events will succeed as:

$$R^{\mathrm{N}}$$

that is, R raised to the Nth power. For example, if R was .90, and N was 3, the probability of the 3-chain sequence would be .729.

Let us reverse the problem somewhat, and examine the characteristic length of a logical chain that will succeed before breakdown occurs, given the value of R. This is called the expected average value of N, written as $E(N)$. The formula turns out to be:

$$E(N)=\frac{1}{1-R}$$

There are some very interesting behavioural consequences implied by the above formula. We should note that the value of R is likely to be quite high in absolute terms, and it is interesting to see the effect that small differences in the value will make.

Suppose that we have three individuals, which we will call 'A', 'B', and 'C', each having a different characteristic value of R. Individual 'A' has an R value of .9900, individual 'B' a value of .9990, and individual 'C' the highest value of all, .9999.

Substituting the values of R into the above formula, we can work out that the average expected values of N, $E(N)$, are 100 for individual 'A', 1000 for individual 'B', and a full 10000 for individual 'C'!

Converting the above expected values into time, we use the value of 230 ms (our assumed average human pulse train length) for each of the N units. Individual 'A' has an expected mean time before failure (MTBF) of 23 s. Individual 'B' has a MTBF of 230 s, or just under 4 min. Individual 'C' has a MTBF of 2300 s, or about 38 min.

For many kinds of everyday activity, we humans do not require very long thought chains. If a particular thought chain was one that required, say, only ten steps for completion, there would be very little absolute difference between the values of R to the N in the three individuals mentioned above. Individual 'A', with the R of .9900, would have a probability of completing a ten-link chain of .904. Individual 'B', with an R value of .9990, has a final probability of .99 of completing the chain, and individual 'C' ends up with a .999 probability.

If we had some sort of IQ test which consisted of large numbers of easy items, each of which could be solved using programs of ten links or less, and these programs were known to our three individuals, it would take a fairly large number of test items to be administered before we could reliably identify individual 'A' from the others. A 100-item test, with each item requiring ten pulse train links, and a time limit placed on each item, should result with 'A' having a score of 90 on the test, 'B' with a score of 99, and 'C' with a score of 100. We would not be very confident on the basis of these results that 'B' and 'C' were really different, and not entirely certain (taking into account normal sampling variability) that 'C' was really better than 'B'.

Now let us step up the difficulty by a factor of 100, and set a test consisting of items with 1000 links each. Each item requires just under 4 min of thinking to arrive at the correct answer. Individual 'A' breaks down on this test completely. He has a probability of solving any single item of only .00004 – lucky guesses excluded. Individual 'B' has a respectable probability of solving any item of .368. Individual 'C' still finds the items fairly easy, with a probability of .904 of getting the answer to an item in the alloted time. If we were to administer a timed test with ten such items to these three individuals, we would expect scores of 0, 4, and 9, which spaces the individuals out nicely. We would therefore be fairly confident that we have distinguished between them with a small number of items (noting, however, that the ten items take longer to administer, as a practical point).

Our model mirrors the commonplace observation that people of very high intelligence (as inferred from their jobs or accomplishments) are not vastly superior in everyday tasks to people of ordinary intelligence. The model also fits in with the feeling of many teachers that it is virtually impossible to teach very difficult tasks to some people, no matter how long one persists (for example, higher mathematics to ESN children).

Reception Probability and Tests of Memory

The above account has so far talked about intelligence tests that might distinguish between individuals with differing levels of R in terms of the presentation of problem solving tasks which require a few seconds or even minutes for the solution of a single item.

Most psychologists will be familiar with the fact that simple tests of vocabulary have been found to be one of the best ways of measuring intelligence, as judged by the criteria implicit in the technique of factor analysis. Vocabulary tests almost always have amongst the highest factor loadings to be found among the various subtests in a typical IQ test.

In a multiple choice situation, it is sometimes possible for a person to use his knowledge of a foreign language, or Latin, to make a good guess at the meaning of an unfamiliar word, but in many cases, a given word presented to the test taker is defined immediately or not at all.

The implication seems to be that tests which primarily sample the amount of stored information are good measures of intelligence, and it is difficult at first to see how we can relate this fact to the previous paradigm. In order to do so, we must come back to the issue of redundancy.

In the introductory sections of this chapter, we discussed how redundancy might be built into a logical system to extend the MTBF. Our own performance on mental tasks is helped by redundancy in exactly the same ways that have been discussed. There is no real doubt that the brain has massive amounts of redundancy built into it, and it could be argued that our mental superiority as a species is simply a function of this redundancy, as it confers very significant performance advantages.

Some replications of learned sequences will come about from simple repetition without any real conscious effort on our parts, whilst others will occur by the conscious application of rote learning. What we may have thought of as 'the strength of the connection' in some simplistic learning theory is in reality the number of replications of a complete logical sequence that we have stored.

To see just how replication can help, let us return to our previous example of the three individuals with differing values of R. Individual 'B' had an R value of .9900, and when we gave him the 1000-link test item, he had a probability of only .36769 in getting it correct. Let us now assume that 'B' had only a single replication of the logical chain at that time, but we have now given him additional training in the problem, and he now has a second full replication of the chain. We now present him with another similar test item, and he 'sets to work', setting off both of his logical chains at once. Now we have to take into account that even if one or other of the chains fails, the other might not. We assume, of course, that the probabilities are independent. In the first instance, the probability of failure was $(1-R)$, or .63231, but the addition of the second chain means that the above probability is raised to the power of 2, or squared. Turning this around, we find that the probability of getting the test answer correct when 'B' has two chains goes up from .3677 to .6002, a very useful gain. If he studies some more, and eventually gets four replications, his probability jumps to .8402, and eight full replications would bring it up to .9745. Thus, we see the value of 'learning', or the creation of these replications. Individual 'B', given enough practice, can outperform individual 'C', provided that individual 'C' does not acquire additional replications as well.

Most adult humans engaged in the 'mental' sorts of jobs, such as being professors or students or clerks, will have a large part of their total memory capacity used up in storing replications for performing various kinds of tasks. However, we do not differ so much from individual to individual in terms of our brain size; the ratio of smallest to largest in the normal adult is perhaps 1 to 2, with the average male human brain being about 1.5 litres. Our memory molecules and the PSDs will be exactly the same size from one individual to the next, and the only way that it would be possible for one individual to have a significantly greater amount of memory capacity would be to have significantly larger numbers of synapses, or, possibly, more PSD hexagonal cells at each synapse. There is no evidence known to us (one way or the other) which suggests that we might differ from each other at this level, but we personally feel it unlikely that such differences would occur. It is not necessary, however, to make such presuppositions to explain our demonstrable differences in memory capacity.

Only one additional assumption is required at this stage, which is that our memory store is finite, and fills up to capacity at some relatively early point in our lives.

If we accepted this assumption, it poses the question, 'How can we continue to learn when our memory storage capacity is used up?' The answer must be that we then make room for new 'facts' by losing old ones. The loss of memory is hard to dispute, but the famous work of Wilder Penfield many years ago that demonstrated (by direct electrical stimulation of the cerebral cortex) that very old memories could be vividly recalled when it was 'thought they had been lost' has given rise to the popular canard that our memory is infinite. Were that it were so! The indisputable fact is that we seem to be unable to recall information when we wish to. Also beyond doubt is the fact that all molecules have limited lives. The so-called half-life of RNA is measured in days, and whilst eRNA might turn out to be longer lived than other RNA species it too must have a finite half-life. Each time we lose a molecule of eRNA, we are losing some of our stored memory. What cannot be dismissed out of hand is the possibility that our memory capacity is such that at normal rates of acquisition, we would not fill up for a very long time – say, 90 years or so. Again, that seems to be improbable, as full brain size is reached very early in life, and there is little evidence for change beyond the age of 18 except in the negative direction.

We believe, on the basis of what poor evidence does exist, that the human brain is full by the mid-twenties, and that the further acquisition of new knowledge beyond that point is at the expense of other memory.

Of course, memory would not be lost in any systematic way, and (unfortunately!) there seems to be no way to control what is lost to make room for the new. The rate of memory loss can probably be controlled in an odd sort of way, because it is a function of acquisition rate. That is, the more we learn (after we are full) the more we lose.

If our reasoning is correct, it implies that people past their mid-twenties who are still actively involved in the acquisition of new knowledge will find themselves beset with the problem of the rapid loss of information that was thought to be safely tucked away somewhere. This may explain the 'absent-minded professor' syndrome.

Now let us return to the question of vocabulary tests. If individuals do have the same basic memory storage size, and a lot of this is used up by redundant storage of logical chains that are formed to give us performance advantages (perhaps of necessity), it follows that the individuals who can be demonstrated to have more facts stored away (as measured by the sampling technique of IQ test items) must have been able to store them by virtue of the fact that they did not use up so much of their memory for redundant storage.

Who are the individuals who have not used up so much of their memory capacity for redundant storage? It must be the individuals who do not need to, because they already have a high probability of task success. Of course, those are the individuals with high levels of *R*! In other words, high *R* means that you do not need to have as much redundancy as somebody with a lower level of *R*. The smaller amount of redundancy means that additional memory capacity is available, and so the high *R* individual ends up with a larger vocabulary, and a greater number of facts and figures at his mental fingertips. When we now sample the stored knowledge with appropriate tests, and also measure problem solving ability, we find the same individuals high in both. Thus, we have a high correlation between fluid and crystallized intelligence, which combine to give us our familiar global '*g*'.

Sex Differences in Intelligence

Many IQ tests show that there are no significant differences in the mean, or average scores of IQ tests between men and women. When a new IQ test is being constructed, it is quite easy to find test items that are not answered correctly by the same proportions of men and women, but these are usually then eliminated from the test. So, to some extent, the fact that no differences are to be found in our current IQ tests between the averages for the two sexes is hardly surprising.

There are, however, differences in the standard deviations or variability of the scores of the two sexes in many IQ tests, with men showing the greater variation. The distribution of many IQ test scores are such that they approximate to the normal (Gaussian) curve. If we regard some arbitrary lower score on a test as a classification like "educational subnormal" (ESN) and some equally arbitrary high score as "gifted", we then find that there is a higher ratio of ESN males to women and likewise we find that there are more gifted men than women, even though the average scores are the same.

The brains of men and women are known to differ somewhat in size, with women having about 200 g less tissue. On the other hand, they also tend to have less body mass, and over the phylogenetic scale, the best index of species intelligence is the ratio of body mass to brain size.

Do men and women differ in terms of the *R* parameter, either in terms of the mean or in variability? If so, what may possibly account for it?

The variability question is in reality two questions, as the reason for the higher proportion of gifted men does not have to be the explanation for the higher proportion of ESN men.

Looking at the question of the upper tail of the IQ distributions first, we must consider what it means in terms of tasks to have a very high value of R. Our previous calculations have shown that the superior performance of high R people only becomes noticable when the task length increases. People with the very highest levels of R can maintain thought processes with a time course of hours, and perhaps longer. Tasks which require these levels of concentrated mental effort are usually to be found in the realms of mathematics, science, and technology of various kinds, although a case can be made to include the upper echelons of most professions. It is at these upper levels, of course, that we find an inbalance in the sex ratios. Are there more male chess grand masters, for example, because there is a shortage of women who have the necessary level of R, or are there other explanations for the ratio differences?

If there was a genuine superiority of the male brain, it should be manifested as a higher mean level of R, which in turn should show up in our conventional IQ test norms. As we do not find this to be the case, it leads one to suspect that an additional factor must be posited to explain the shortage of women from even those activities where there seems to be little evidence suggesting social discrimination.

We believe that there is such a factor, which has the effect of invalidating high levels of R in women. Put another way, there are probably just as many women who could concentrate on a chess problem for hours on end as there are men, but there is something that prevents them from doing so. We believe that the additional factor is one best explained using a computer analogy.

Computers are constructed so that they can attend to many things at the same time. However, unlike the brain (which is extensively parallel) computers often have only one central processing unit. When a computer is carrying out a computation task, an event might occur which requires immediate attention. In this case, the computer hardware generates an interrupt condition, and the software routines save the context of the job that the computer was working on. The computer then turns its attention to the interrupt condition, takes care of it, and then restores the context of the interrupted task and resumes computation.

Problems occur in computer systems if there are too many interrupts in too short a space of time. Each interrupt puts a demand on certain dynamic memory capacity the computer has available. If there are too many interrupts (which can be nested, with a higher level interrupt interrupting an interrupt, etc.) the computer system might run out of dynamic memory. If that happens, the system usually misbehaves, or at the very least suffers the temporary loss of data and wastes some time.

We suggest that adult women have a far larger number of interrupt conditions to attend to than men. These interrupt conditions are stimuli which are attended to because there is eRNA in the brain scanning for the condition. This particular eRNA is thought to be there not as a result of learning, but comes from DNA transcription. The biological purpose of such eRNA is to program the woman to respond to such things as an infant's cry, or a possible threat from some external source. Men have such eRNA as well, of course, and all that we are suggesting is that women have a much larger repertoire of events that they are biologically programmed to respond to.

When some of these events occur, we suggest that it may be impossible to effectively 'store the context.' Chess problems, as an example, quickly get into the combinatorial explosion of having tens of thousands of possible moves, looking ahead only a few turns. If an interrupt of sufficient magnitude occurs whilst solving a difficult chess problem, the only way that one can return to the problem is to start at the beginning.

If, before you have reconstructed the point at which the interrupt occurred, another interrupt occurs, the effect can be very frustrating. The task enjoyment is effectively destroyed, and the person is unlikely to persist in the activity. This, we believe, is the reason for the lack of women in the professions which require long periods of concentration.

In some support of this interpretation, we note that some activities, notably literature, have 'hard copy output' as an inherent part of the activity. In such cases, it is usually less frustrating to get back into the problem after an interrupt because one can refer to ones own output to re-establish the context. Thus, we find that there are large numbers of female authors, etc.

We feel that tasks such as musical composition which might at first sight seem to be in the same category (but which also lack the same proportions of women) will also be found to be tasks which require the extended periods of concentration. Reputedly, composers often compose a piece 'in their head' and the hard copy output stage occurs at the very end of the creative process. A literary work is probably composed more in outline at the onset, with the detail provided at the time of writing.

The Measurement of Biological Intelligence

In this section, we consider how we might be able to measure biological intelligence. The previous sections have discussed how and why fluid and crystallized intelligence are related, and it follows that our tried and true (albeit highly criticized) standard IQ tests will correlate well with biological intelligence.

The criticism that everybody will accept of such tests is in the applicability of a given test to a given sample of subjects. We have mentioned how the number of stored facts is measured by means of sampling, and everybody would agree that this is subject to a high degree of measurement error.

Is there, then, a way of way of directly getting at some estimate of the value of *R*, the recognition probability, without the constraint that any culturally related test items would introduce?

The Electroencephalogram and Intelligence Measures

Attempts to find 'culture free' tests of intelligence go back a very long way, and other chapters in this book mention some early studies in this area. In the last 2 decades, a number of studies have appeared reporting attempts to relate intelligence scores to measures derived from electroencephalography, or EEG. EEG is the measurement of the tiny electrical currents of the brain. These can be measured by the insertion of electrodes directly into brain tissue, or rather more weakly, by means of electrodes placed directly upon the scalp.

Many of the studies which have attempted to relate EEG to IQ (for reviews, see Ellingson 1966, Vogel and Broverman 1964, Shucard and Horn 1972, DE Hendrickson 1972, Callaway 1975) have used the technique of the averaged evoked potential (AEP).

The AEP technique consists of presenting a subject with a repetitive stimulus, and recording the EEG signal at the same time. A record is made such that the point of stimulus onset can be determined in the EEG recording, when a predetermined number of stimuli have been presented, the EEG records are then processed so that a given time epoch from stimulus onset is averaged over the total number of stimulus presentations. This can be done in a number of ways, ranging from the original method of superimposition of the signals on a storage tube, to the more modern analog-to-digital (A/D) conversion methods which use computer technology. The end result is a wave form which is the 'pure' response of the brain to the stimulus, as (hopefully) the other activity that was going on at the same time each stimulus was presented was uncorrelated, and was thereby cancelled out. The technique is known more generally as 'signal averaging'.

Various of the studies mentioned in the above reviews were able to find significant correlations between IQ and measures derived from the AEP processed records (for example, see Ertl and Schafer 1969).

One of the most common of the scoring techniques, given the AEP, wave form, was to scan along the record from stimulus onset, and mark the peaks and troughs of the wave form. The distance from stimulus onset to each peak or trough was noted. The distance on the horizontal axis of the graph corresponds to time.

The assumption was made by these investigators that the peaks and troughs they noted in each record were, in some sense, records of the same neural event from one person to the next. That is, peak 3 (say) for subject 'A' represented the same thing as peak 3 for subject 'B', even though they would not in general be found at exactly the same point in the record.

When the times (latencies) for these various peaks and troughs were correlated with IQ scores, it was found by some of the investigators that there was a significant negative correlation between IQ and the time measures. High IQ subjects tended to have their peaks and troughs occurring earlier in the AEP record.

The correlations reported tended to be in the range −.30 to −.45 or so, which were unlikely to arise by chance for the various sample sizes used. At the same time, a number of other studies appeared to show zero order and non-significant correlations. Evidently, whatever was being measured by these studies was subject to variability of technique, or sample, or both.

In order to relate the findings that are referred to above to the theory presented here, it is necessary to consider in some more detail what the EEG is, and what exactly it is measuring. At one level, we can say that we are simply recording the electrical activity of the brain in a rather gross way. However, we need to know what that electrical activity represents in order to interpret it.

The subject of the electrogenesis of the EEG is something which has concerned a number of investigators ever since Hans Berger (1929) discovered that he could measure electrical potentials directly from the scalp. After some decades of experiment and debate, there still does not seem to be a clear-cut consensus on the meaning of the EEG, and what exactly is causing the electrical currents.

One study that we personally find both illuminating and convincing is that of Fox and O'Brien (1965). Fox and O'Brien recorded the response of single neurons in the cortex of experimental animals to a visual stimulus. Large numbers of presentations of the stimulus (usually over 4000) were made, and the pulse train activity of the neuron immediately after the stimulus presentation was recorded. The elapsed time from the onset of the stimulus until the occurrence of each pulse (within a specified epoch) was measured, and a histogram of the numbers of pulse occurrences with specified ranges of times (bins) from stimulus onset was made. After this series of presentations was complete, the microelectrode used to record the activity of the single neuron was pushed a bit deeper into the cortex until it was not within a single axon, but in the surrounding material. The electrode was now able to make a direct internal record of the EEG currents within the cortex at that point. During this second recording, a much smaller number of presentations of the same stimulus were made.

Fox and O'Brien collected such data on nearly 200 single units. They analysed their data by visually comparing the histograms with the AEP waveforms from the same point. Figure 16 shows two of these comparisons.

In Fig. 16, there is a remarkable correspondence between the envelope of the histogram and the AEP wave form; if superimposed they line up almost exactly if

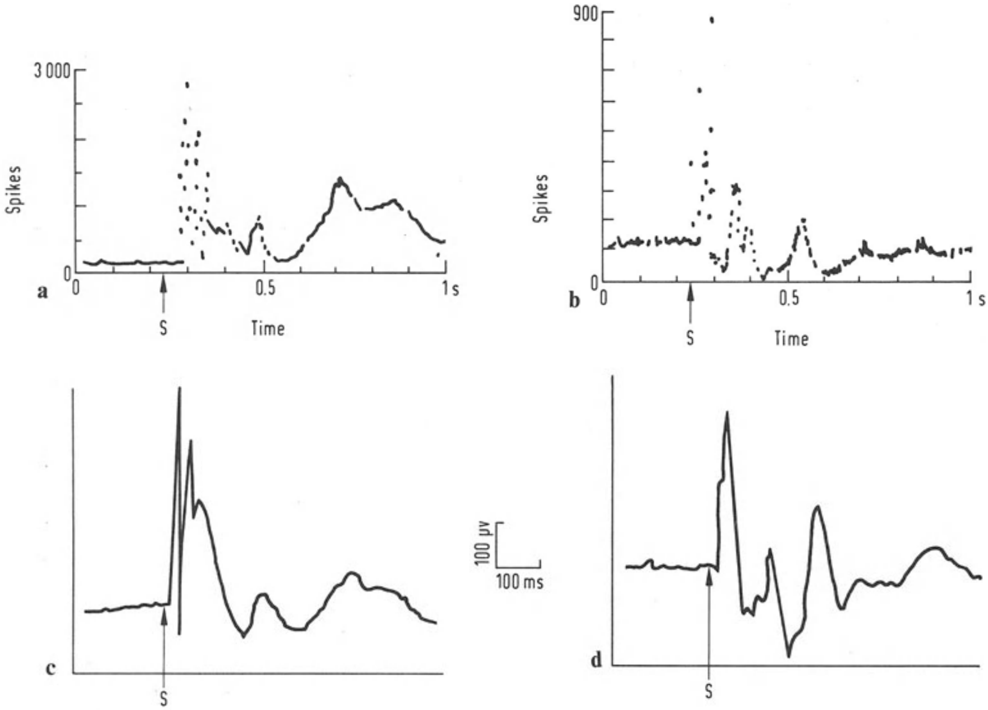


Fig. 16 a–d. The relation between probability of a single cell firing and evoked potential waveform. **a** Frequency distribution of spikes from a single cell in the visual cortex of a cat after stimulation with 4918 flashes; **b** Averaged evoked potential (200 oscilloscope sweeps) recorded from the same microelectrode, after cell death ($r = .60$, $P < .001$). Similarly, spike distribution for a single cell is shown in **c** (3150 sweeps) and the corresponding averaged evoked potential in **d** (150 sweeps) ($r = .51$; $P < .001$). *Ordinate* (for unit distributions); number of times the cell fired in response to light flash. *Abscissa* (for unit distributions); time, in 100-ms divisions. (Fox and O'Brien 1965)

the vertical scales are made to match. The examples shown are said to be chosen from a large number of such close alignments which occurred amongst the 200 neurons that were studied.

The implication of the Fox and O'Brien study and similar studies which have shown more or less the same thing (Vaughan 1969, Creutzfeldt et al. 1969) is that there are a fairly large number of cerebral cortex axonal fibres which are carrying pulse trains which have originated from the primary sensory units (such as the hair cells of the cochlea). These are probably long 'transverse' fibres with thousands of synaptic contacts which allow the functional associative units to be aware of the primary input stimuli.

The correspondence between the histograms and the AEP waveforms, then, we interpret to mean that the electrogenesis of the EEG is derived from the summation of the individual pulses, provided that some pulse trains are predominant in the sense that there are many more replications of them at the same time and place. If all the pulse trains going on at any one time within a confined area were different, the result would probably be that little or no EEG activity would be recorded (this is because it is momentary voltage changes, or differences, which are recorded, as opposed to some absolute sum of electrical activity.) If we think back to our early example of human speech, we noted that the larynx imposes a pattern on the higher frequencies.

It is probable that many of the individual hair cells firing in response to a phoneme are sending out very similar pulse trains, tracking the amplitude changes from moment to moment. These coordinated pulse trains summate to give rise to a fairly easily detected AEP response.

If our interpretation of the studies cited above is correct, it means that we can interpret the EEG AEP waveform as a kind of picture of the individual pulse trains that were set off by the primary stimulus. What follows then is the possibility of fairly direct measurement of the amount of error in the pulse train transmission, which might in turn have a monotonic relationship to the *R* parameter.

The Effect of Transmission Errors on Waveform Appearance

In order to be able to understand the nature of the EEG waveform and its relationship to the pulse trains which might underlie it, we carried out a series of computer simulation studies.

A computer program was written which generated a series of real numbers, with each number in the series being a specified constant amount greater than the preceding number. The series was intended to represent the time elapsed from occurrence of the first pulse in a pulse train until the arrival of the subsequent pulses in the train. Thus, a typical series might be 8.0, 16.0, 24.0, 32.0, etc., with an implied unit of milliseconds. As the series was formed, by repeatedly adding the constant to a developing sum, a random number was generated from a distribution with a mean of zero, and a specified standard deviation, and added to the constant. Thus, depending on the chosen standard deviation of the random number series, the series might actually read 7.96, 15.84, 24.09, 32.01, etc., rather than the perfect (error free) series mentioned above. The intent of the addition of the random term was to simulate the error in temporal intervals between one pulse and the next as they travelled down the axon. Our algorithm was such that the error in our simulated pulse trains was cumulative, in that the location of the second pulse was a function of the error added to the first pulse as well as the second, the third pulse location was a function of the first three random numbers, etc. The position of each successive pulse in the train was therefore more indeterminate than the preceding pulse. The program could be made to generate any number of these pulse trains, each with a different series of random numbers drawn from the same distribution, which was specified at the onset of the program. The final output of the program was to generate a histogram of the pulse interval positions. In effect, our program created data similar to the data collected by Fox and O'Brien.

One difference between our program and real data is the position of the very first pulse. As we always start at a zero time base, the first pulse in the train has very little error.

Figures 17–20 are a selection of computer drawn histograms produced by our program. In this series, we generated pulses 24 ms apart in order to show more clearly what happens when the random number parameter is varied. The histograms, however, have to be interpreted with some caution, as they are not drawn to the same scale. The vertical axis on each histogram in the series changes, to preserve the amplitude detail. Each histogram represents the summation of 500 generated pulse trains.

The first histogram (Fig. 17) in the series is the trivial case of no error; each pulse is in exactly the same position as the corresponding pulses of the other trains.

Figure 18 shows the series with the first amount of error added to the system, in this case a random number with zero mean and a standard deviation of 0.2. Note the change in the vertical scale (the full range is 512 plotting units), which has dropped from .98 to 1.83. This histogram clearly shows the effect of the cumulative error, as the height of each pulse drops and the base

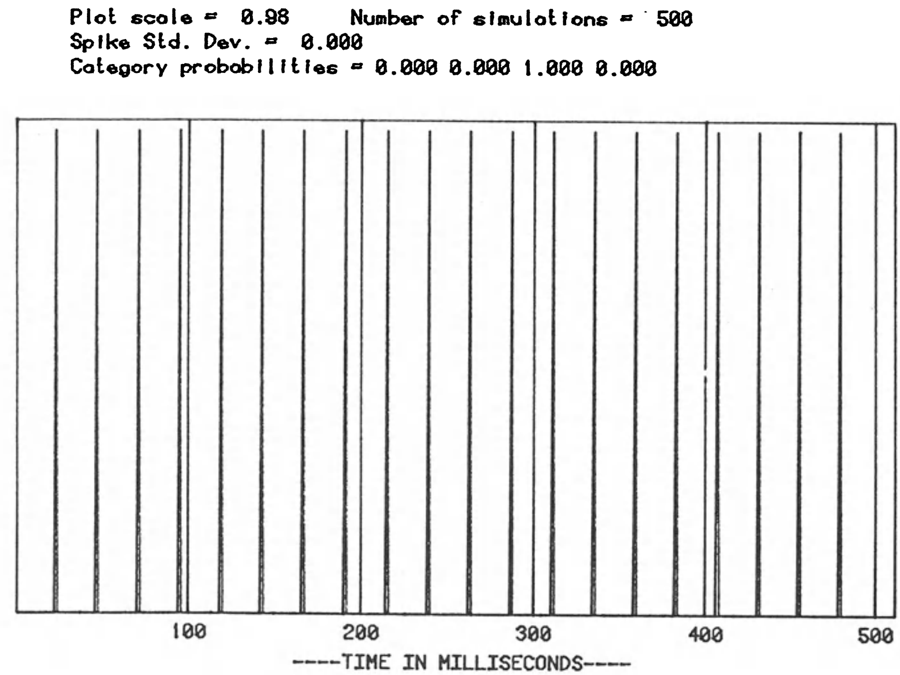


Fig. 17. Trivial case of pulse train histogram with no error added, showing exact location of each pulse

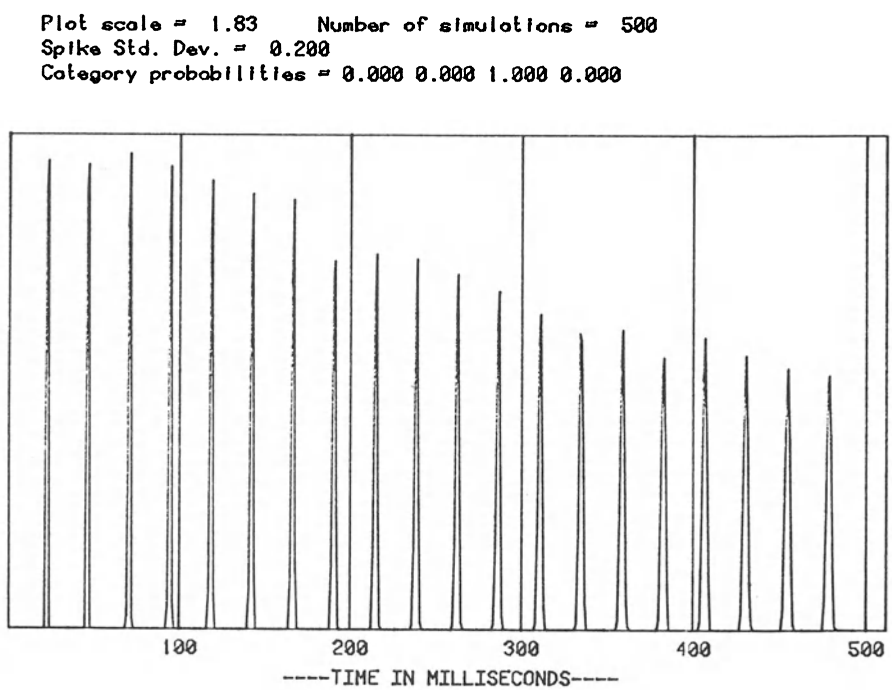


Fig. 18. Pulse train histogram, with standard deviation of 0.2. The cumulative error makes the right-most modes shorter and broader. Plot scale is 1.83 units

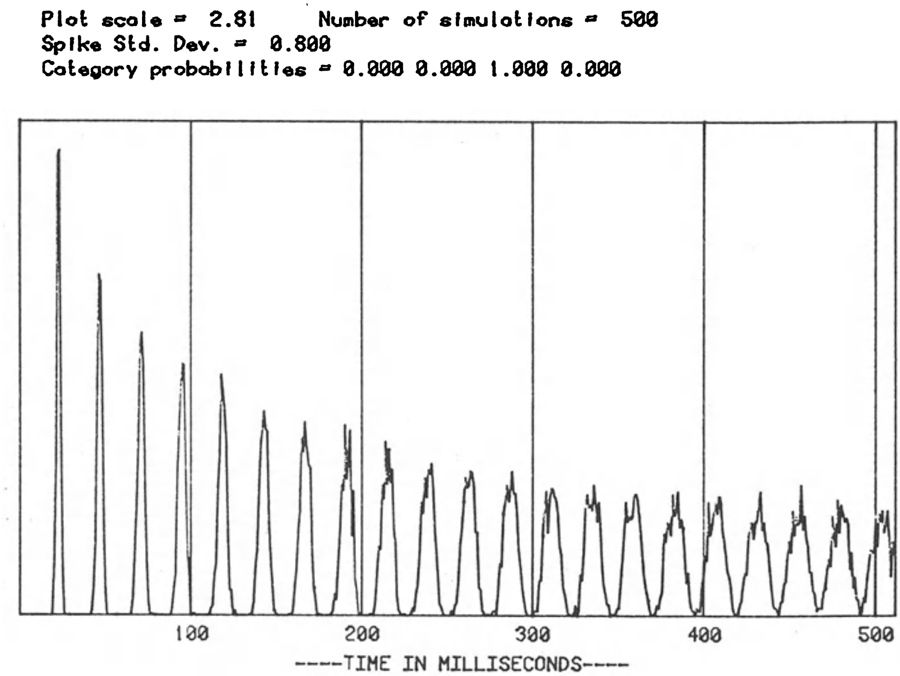


Fig. 19. Pulse train histogram, with standard deviation of 0.8. The bases of individual pulse modes are now touching at the base of the histogram. Plot scale is 2.81 units

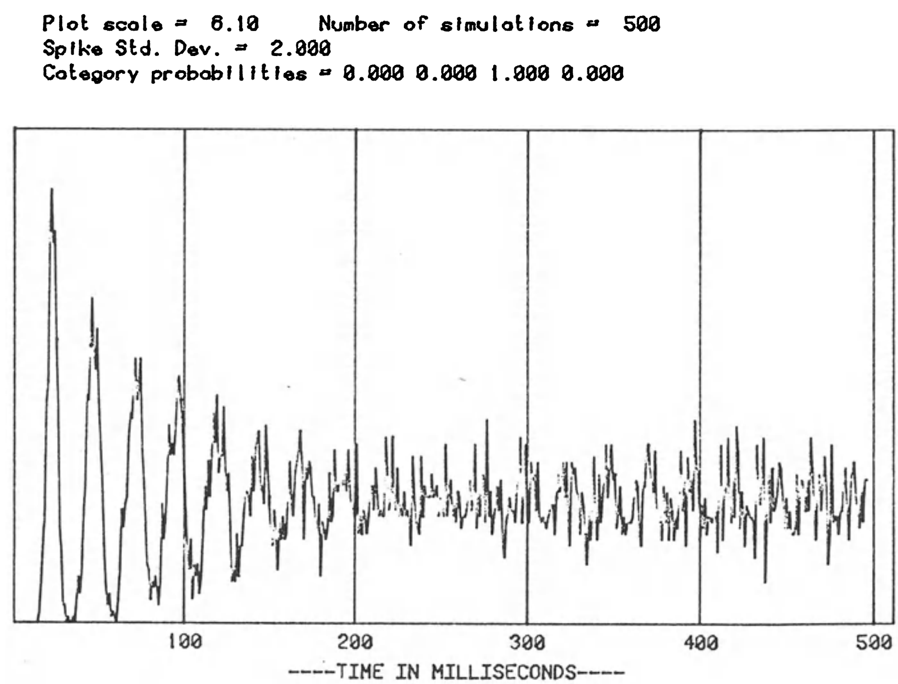


Fig. 20. Pulse train histogram, with standard deviation of 2.0. Individual pulse modes have now disappeared in the right half of the histogram. Plot scale is 6.10

enlarges as we go from left to right into the record.

Figure 19 shows the standard deviation with a value of .8. The position of the original pulses can still be clearly seen as a mode, but the bases of each mode are beginning to overlap.

Figure 20 has a very large amount of error added, with a standard deviation of 2.0. You can now determine the original position of just the first seven pulses, with a suggestion for the eighth. Beyond that, the record is just a spiky line.

If we were to replot the last histogram with the original plot scale, the spiky nature of the record would smooth out, and the picture would begin to look very much like the wave forms we generate from AEP recordings.

A New Measure of Intelligence

We have briefly described how AEP records have previously been scored to find a measure that was correlated with IQ.

Our simulation study helped us understand why some investigators were able to obtain significant negative correlations between the locations of peaks and troughs and standard IQ scores.

Our reasoning proceeds as follows. If high-IQ people have high levels of R, it may be because their axonal pulse train transmission has less error in it. If axonal pulse trains give rise to the AEP waveform, then we should be able to see differences between the AEP records of high-IQ people and low-IQ people. If the error is cumulative, as described in our simulation study, then what we should see is the AEP record getting smoother for the low-IQ people the farther we go into a record. Finally, if the record is getting smoother, it is likely that tiny peaks and troughs are completely 'smoothed out' and merge with the next peak or trough. If that is so, then as we count the peaks and troughs as we go into the record from left to right, we have to go further into the record to get to, say, peak number 5 if there is a lot of error, as compared to a record where there is very little error.

Put another way, we are saying that the research workers who used the 'traditional' way of measuring peak and trough location were simply not comparing like with like. Nonetheless, what they were doing had a crude measure of validity, although the investigators tended to give odd names like 'speed' or 'reactivity', etc., to their measures.

We were convinced that a better measure of the error could be found than the peak and trough counting method, which was rather subjective at the best of times. In looking at the records, we noticed that as the waveforms of the low IQ records became smoother, the circumference of the waveform envelope became shorter. If we thought of the waveform as a piece of string, and we went a standard length into the record, cut the string at that point, and pulled it straight, the high-IQ people would have longer waveform strings than the low-IQ people.

Having noticed this, we resolved to try it. Figure 21 shows some data that were published by Ertl and Schafer (1969). They represent the AEP records of ten selected high-IQ subjects and ten selected low-IQ subjects.

The records shown in Fig. 21 were measured by the simple means of using pins and thread, after a photocopy was made of the original figure. The thread was laid over the waveform lines, held in place by pins, and then cut, pulled out, and measured. As Ertl and Schafer had conveniently recorded the actual WISC IQ scores of each subject, we were able to compute the product-moment correlation between the string lengths and the published IQ scores. Our result was a correlation of .77. Although this result was possibly inflated by the fact that the subjects selected by Ertl and Schafer were in the IQ distribution tails, we were sufficiently encouraged by the result to try the measure on data of our own. These results are reported in the next chapter.

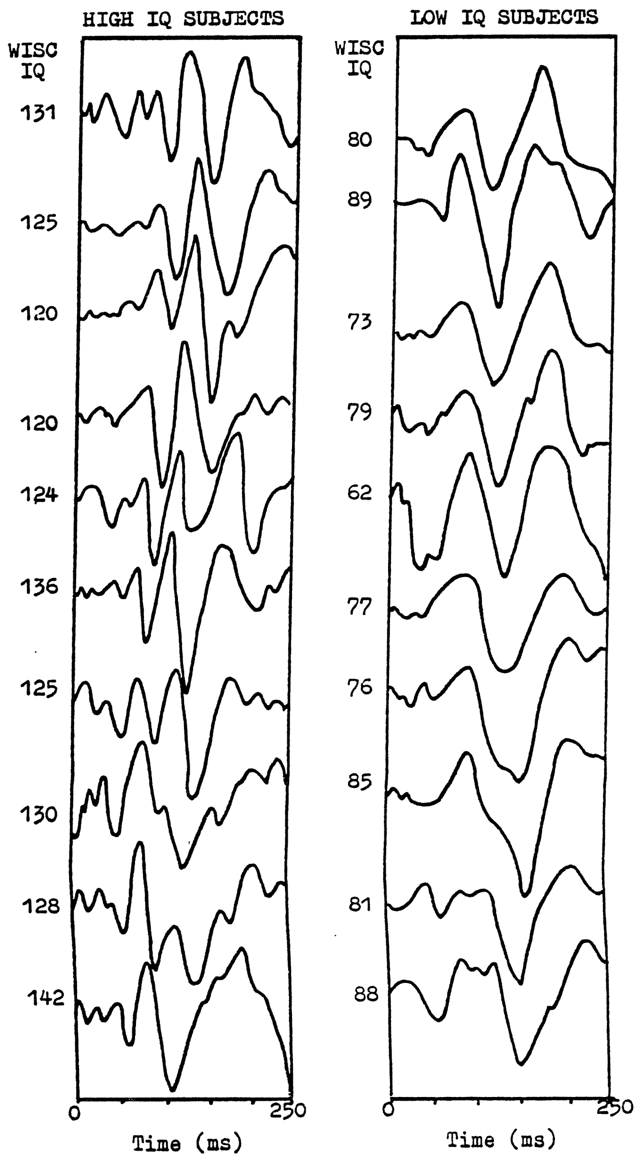


Fig. 21. Evoked potential waveforms for ten high- and ten low-IO subjects. The individual WISC IO scores are shown to the left of the beginning of each waveform. (Ertl and Schafer 1969)

Error and Other Measures of Biological Intelligence

Other chapters in this book record the relationship of standard IQ measures to such things as inspection time (Brand and Deary, this book), and choice reaction time (Jensen this book). The next chapter (DE Hendrickson) discusses another statistical measure derived from AEP scores. We feel that all of these other measures can be related to the paradigm described in this chapter. However, space limitations prevent the detailed discussion that would be required to present a convincing argument in respect to these other measures. We will leave the results presented in the next chapter to speak for themselves.

7 The Biological Basis of Intelligence. Part II: Measurement

D.E. Hendrickson

Introduction

The preceding chapter has presented a detailed theory of the biological basis of intelligence. This chapter reports a major research study that was carried out in an attempt to verify some of the specific predictions made by the theory.

We have previously carried out research in using the EEG as a means of measuring intelligence. Our first study (DE Hendrickson 1972) successfully replicated findings previously reported by Ertl and Schafer (1969). A summary of our replication study has also been reported in a recent paper (Hendrickson and Hendrickson 1980).

Subsequent to this first study, the theoretical work on the model of intelligence was extended, and the 'string' measure described in the previous chapter was formulated. Unfortunately, the magnetic tapes containing the raw EEG recordings for the previous study had been reused, and it was not possible to try the new measure on any existing data of our own. An attempt to use the measure on the published EEG waveforms in the Ertl and Schafer paper gave us a correlation of .77 with WISC IQ scores. The correlation was established, however, on only 20 published records that had been selected on the basis of being representative of high- and low-IQ subjects. We were encouraged by this finding, and resolved to try the new measure on a set of data gathered for that specific purpose.[1]

1 The raw data from the present study have been retained in both analog and digital forms, together with test protocols, etc., and can be made available to other researchers by arrangement.

During the collection of data for the present study, we were given access to a group of 37 psychology students at the Hatfield Polytechnic. Data were obtained from this sample using the techniques described here, but our analysis procedures had not been completely established then. Our string measure, based on an edited selection of 32 records taken from the full testing session of 100 records, was computed. These scores were turned over to Dr. S. Blinkhorn at the Hatfield Polytechnic, who then compared them with the student's scores on the Raven's Advanced Matrices (RAM) test, which had been used for selection of the student sample. Despite the restricted variance thus built into the IQ measure, Blinkhorn found that our tentative string measure and the RAM scores correlated to the extent of .47, which, when corrected for the estimated attenuation, gave a 'corrected' result of .80. These results have been reported in detail elsewhere (Blinkhorn and Hendrickson, 1982).

For the present study, we wanted a sample that included a full range of IQ scores, but at the same time was homogeneous in age and in general cultural terms. Accordingly, we decided to use a group of school children drawn from the Greater London area. Schools were selected on the basis of neighbourhood within this area to ensure a measure of heterogeneity with respect to 'social class.'

As well as trying the 'string' measure on a reasonable sample, we wanted to carry out detailed analyses of the data to see if other EEG measures of IQ could be found.

Finally, we wanted to establish standards of procedure and scoring that might allow our methods to eventually move from the laboratory to clinical settings. We accordingly report our techniques in some detail, together with a brief history of some of the problems we encountered during our research.

Experimental Design and Method

Subjects

Our primary subject sample consisted of older schoolchildren or children in the same age range. All of these children lived in Southeast England. The average age of this group was 15.6 years (S.D. = 1.13).

The primary sample was not random, but an attempt was made to draw the subjects from various sources, in order to have a reasonable cross section of social backgrounds. Although not all of the children were attending school, we will refer to this primary sample as our 'school' sample.

There were a total of 219 children in the school sample. Boys comprised 121 of our main sample, and 98 were girls.

Of these, a total of 122 were drawn from four main sources in the Camberwell (Southeast London) district, which is a predominately 'working class' neighbourhood, but which also includes a small 'upper middle' neighbourhood called Dulwich. The four sources of subjects were two schools, and two social clubs with large memberships. Subjects drawn from this sample were offered payment of three pounds sterling for participating in the experiment, and the great majority of them accepted the payment. All of the testing for this group was carried out at the Institute of Psychiatry, University of London, which is situated in the centre of this district.

Another small group called 'other' consisted of 18 children drawn from the same area, who were contacted by a variety of methods. Some of them were working on a part-time basis as porters in the Institute. They tended, however, to come from the Dulwich area mentioned above. Like the group above, the 'others' were offered payment of three pounds, and were tested at the Institute of Psychiatry.

The other school group consisted of 79 children drawn from the Southwest London area, just inside the Surrey border. Children were obtained from a state comprehensive school, located in a middle to upper middle class neighbourhood. None of the children from the Surrey school were offered payment, and all of them were tested in facilities provided by their own school.

In all cases where children were obtained through schools, it was with the permission of the school officials and the local education authority. In addition, children were given letters to take home to their parents explaining the research and asking their permission for their children to take part in the research. There were more than adequate numbers of volunteers, and no selection factor was used to choose amongst those who did agree to participate other than 'first come, first taken'.

In addition to our main sample described above, we had two special purpose subsamples.

One of our subsamples consisted of 19 volunteer subjects from an international society called Mensa. Membership in this organization is conditional upon taking an IQ test administered by the society and obtaining a minimum high qualifying score on the test. We were interested in having members from this group precisely because of their membership conditions, as it gave us a group known to have a high level of IQ as measured by conventional means. The Mensa group traveled to the Institute of Psychiatry, and were offered reimbursement of their travel expenses. Some of the volunteers refused payment on the grounds that they wanted to help our research project. There were 12 men and 7 women in the Mensa group, with an average age of 28.7 years, and an S.D. of 6.54.

The second special subsample consisted of 16 court stenographers. Of these, 15 were women. The average age of this sample was 42.4 years, with an S.D. of 9.57. This group was obtained because they happened to approach us during the research and asked to have a psychological assessment carried out on some of their membership which they hoped would be of some use in negotiating better terms of employment. It was agreed that we would write a short report giving the mean IQ scores as measured by our conventional tests which could be used by the group in their salary negotiations. No payments were made to this group.

The data from approximately 25 people tested were rejected completely. In almost all cases, this was because of excessive 50 cycle interference with the EEG recordings. In the majority of these cases, the test session was not completed, but several were carried through to the end and rejected after examination of the recordings. Two Mensa subjects were lost, and all of the others were from the school samples. One young man was tested on two different occasions, and was found to have, or produce, an excessive amount of 50 cycle interference on each occasion.

A final group of 15 subjects drawn from a state institution for the 'severely subnormal' was also tested, in the sense that an attempt was made to obtain EEG recordings from them. These data are not included in this report in detail. Further comments about this subsample are made in the discussion.

EEG-Related Techniques and Procedures

Subjects were seated in a recliner chair with full head and neck support and asked to keep their eyes closed throughout the stimulus presentation. The special testing room at the Institute of Psychiatry was sound deadened, but the room used in the Surrey school was not. A recliner chair was not available in the Surrey school, so the most comfortable chair available was used, with head and neck support being provided by generous numbers of pillows. The rooms were darkened during the stimulus presentation. In the Surrey school, careful note was made of class changing times to avoid the extra noise of bells and children's voices.

Our primary stimulus source was an auditory sine wave generator. This was set to produce a 1,000-Hz tone, with an amplitude of 85 dB delivered to earphones as measured by a sound level meter. The tones were administered for 30 ms, with the sine wave being switched at a zero crossing point to minimize the production of higher order harmonics. The tone was presented binaurally through high quality earphones.

Stimulus presentation was controlled through a programmed device, which varied the interstimulus interval on a pseudorandom basis in the range 1–8 s. Each subject received exactly the same sequence of interstimulus intervals. One hundred presentations of the stimulus were made.

Electrodes used for the EEG recording were silver/silver chloride, which were attached at the vertex and to both mastoids using collodion. Some pretesting experimented with other electrode types and attachment methods but they were all rejected on the grounds of excessive movement artefact. Background EEG was recorded from a bipolar derivation with the active electrode being the Vertex electrode [Cz in the 10–20 system (Jasper 1958)] and the reference being the electrode on the left mastoid. The right mastoid acted as earth. Other bipolar derivations were monitored in initial stages of the research, but as our preliminary examination of the recordings showed the Vertex response as being the most clearly defined, it was decided to limit recording to this position for the majority of the sample.

Subjects were prepared by first cleaning their scalps at the point of electrode attachment with acetone. Abrasion was done with a blunt needle which was filled with electrode jelly.

The electrode leads were interwound to maximize common-mode rejection and taken back to an 'PTT4' EEG amplifier.

This is an amplifier of our own design. Standard amplifiers were rejected partly on the grounds of size and weight. In addition, we wanted to have an amplification system that had no inbuilt upper frequency or notch filter, for reasons mentioned in the discussion. The circuit diagram of this amplifier is included in an appendix to this chapter. The amplifier provided a preset amplification level of 10,000. Amplifier output was fed into one channel of a Yasec CD 1000 instrumentation quality cassette recorder. A second channel of the data recorder was used for marking the onset of the stimulus, with the input signal provided by the stimulus sequence programmer. A 10-μV calibration marker provided a known voltage input signal. This was particularly important because of our suggested scoring method.

Intelligence Testing

All of the subjects with the exception of the Mensa group and the subnormals were tested using the Wechsler Adult Intelligence Scale (WAIS) (Wechsler 1955), as modified for Great Britain by the National Foundation for Educational Research (Saville 1971).

The raw scores obtained from the test sessions were converted to IQ scores using the American norms. Scores were also recorded for the various subtests of the WAIS.

The Mensa group was not tested using the WAIS, but instead each of the 19 volunteers was asked to provide us with their official Mensa IQ score. These scores were obtained from the individuals themselves, and not from the society. The mean IQ scores reported for the Mensa group herein have been adjusted (scaled down) to make them comparable with WAIS IQ scores.

Scoring the EEG Data

The data were collected over an 18-month period, and when approximately one-half of the sample had been obtained, a number of preliminary studies were carried out on the data available at that point.

The EEG data were processed by feeding the signals recorded on the cassette tapes into an AR-11 A/D device attached to a PDP 11 computer. The stimulus marker channel was fed to a trigger on the AR-11 to initiate A/D sampling. The AR-11 converts the input voltages to an accuracy of 10 bits (1 part in 1024).

Various sampling epochs for the purposes of the A/D conversion were tried during our preliminary studies, ranging from 0.5 ms (2.0 kHz) to 10 ms (100 Hz). The waveforms obtained from individual records were plotted and compared visually, in an attempt to judge subjectively the optimal sampling rate. We wanted to record and retain as much of the inherent electrical activity as seemed to be in the signal, but at the same time we had practical limitations in terms of the vast quantities of data that were generated. In the event, we settled for a 1-ms sampling rate for most of the results presented below. A subsample was also processed using a 2-ms sampling rate in addition, for reasons discussed later.

After digital conversion, the digitized data were immediately fed back into the AR-11 into the D/A channels, which were used to drive an oscilloscope. This was used to monitor the process of data conversion, and also provided a means of adjusting the amplification, which was done by matching the output from the 10-μV calibration signal to a constant reference point on the oscilloscope.

Editing of the Data

At the halfway point, we also experimented with a number of methods of editing the EEG data, with the intention of removing records with gross movement artefacts or obvious 50-cycle frequency superimposition. It was our original intention to fully automate this procedure by using the computer to detect 'bad' records, and eliminate them. This, it was felt, would have the advantage of being completely objective.

Various algorithms were tried out on selected sets of data, and we finally settled on a simple one which computed the degree to which a particular stimulus presentation was different from the others. The mean wave form was first established using all available data, and the sum of the squared differences over all data points was then computed as a measure of 'goodness of fit'. Records that were above a given level (a bad fit) were then rejected, and the mean was then recomputed on the basis of the records that were retained.

When a reasonable sample of subjects data had been processed in the way described above, the 'string' measure described below was computed, and correlations were then obtained between the string measure and the IQ measures. We were perplexed to find that the correlations were of the order of zero, or even in the 'wrong' direction, albeit not significantly so.

Further examination of the data revealed the rather interesting fact that the numbers of records retained by our editing procedure varied considerably from one subject to the next. Moreover, it was apparent just from inspection that there was some correlation between this number and the overall IQ of the subject.

Another attempt was made at editing by reverting back to a purely visual inspection of the records, with rejection of those that were felt to show artefact of some kind. This also resulted in low levels of correlations with the IQ measures. Eventually, we discovered that the string measure was very sensitive to the numbers of records included in the averages, and it became apparent that we needed to keep this number constant for all of our subjects.

Accordingly, we modified our editing procedure as follows. The entire session of 100 presentations was digitized and presented momentarily on the oscilloscope. If there were any records which seemed to be grossly distorted, they were noted and subsequently eliminated. The first few records at the onset of the session were removed if they showed evidence of muscle artefact. The number of presentations retained was then kept at a constant 90. Where this required the elimination of additional records (which was usually the case) the records removed were taken from the end of the testing session.

The editing procedure was carried out on tapes which were identified only by number, and the IQ of the subject in question was not known to the experimenter.

The Experimental EEG Scores

The previous chapter has discussed our 'string' measure, and how it was first computed using a piece of thread. We decided to use a PDP11 instead for the present research. In addition to our 'string' measure, we tried a number of other measures, which are described in detail below.

The String Measure. The easiest and least ambiguous way to present our scoring procedure for the string measure is to show the algorithm as it appeared in our FORTRAN program. The vector 'SUM' contains 256 real numbers. Each vector element is the mean of the 90 presentations recorded for the corresponding point.

```
      TEMP=0.0
      D=SUM(1)
      DO 42 J=2,256
      TEMP=TEMP+(D-SUM(J))**2
42    D=SUM(J)
      STRING=TEMP/255.0
```

The final real scalar variable, 'STRING' was used as our string measure. A case could be made for taking the square root of this number (or applying any number of transformations, for that matter) but as yet we have not established the effect of doing so.

The Variance Measure. Our finding that removing the most variable records from each persons sample affected the string measure led us to think that the variability of the individual records, measured in the same way, might be a good measure of IQ in its own right.

Accordingly, we defined another variable, computed in the following way. We first went through the 90 records remaining after our final editing procedure, and obtained the sums and sums of squares of the digitized points. These were held in two vectors, 'SUMS', and 'SUMSQ', respectively. We then computed:

```
      DO 21 J=1,256
21    SUMSQ(J)=((SUMSQ(J)
      -SUM(J)/FN*SUM(J))/FN
      TVAR=0.0
      DO 42 J=1,256
42    TVAR=TVAR+SUMSQ(J)
```

where TVAR was the 'total variance of all points from the central waveform.' FN in the above formula is the number of presentations, which was fixed at 90.0 for the data reported herein.

The Multiple String Score. Another experimental measure was to compute the 'string' measure on each individual record, prior to averaging, and to sum these 90 scores. The formula was identical to the string measure mentioned above, except that the data represented single records rather than averaged data points. The 90 individual string measures were then summed, and the sum divided by 90.

The Zigzag Score. Another measure of variability that we thought might be of interest was inspired to some extent by the original work in this area which had looked at peaks and troughs of the waveform. We wrote a simple procedure to look at the number of times that the waveform changed direction. This was computed on the individual records, rather than the averaged records. A number of experimental runs were made with this measure, and we found that results varied somewhat according to the length of time that we 'looked ahead' to see in what direction the wave was going relative to the current data point. The data presented here (in part) started with points 1 and 2 to establish an initial direction, and then went in steps of 4, commencing with point 3, to see if the direction remained the same. The number of changes was summed, and the sum divided by 90.

The Composite Variance Minus String Score. A 'composite' score, subtracting the string score from the variance score, was computed. By coincidence, the way our data were recorded resulted in the means and standard deviations of the string and variance scores being very close to each other. It was not necessary to convert the raw scores to unit normal form, and the composite score was computed by simple subtraction of raw scores. This score was suggested by Professor H.J. Eysenck after examination of some of our preliminary analyses.

The Epoch

The theory discussed in the previous chapter has stated that the average human pulse train length is about 230 ms. We accordingly used an epoch close to this value; 256 ms.

For comparative purposes, however, a subsample was rescored on a 512-ms epoch. However, the A/D sampling interval was altered to 2 ms for this epoch, which allowed the longer time period to be still represented with 256 data points. This was done for convenience of data processing.

Personality Measures

The Eysenck Personality Questionnaire (EPQ) (Eysenck and Eysenck 1975) was administered to the majority of our subjects. This test instrument provides measures of three aspects of personality, namely, 'extraversion' (E), 'neuroticism' (N), 'psychoticism' (P), and a 'lie' (L) scale.

Test-Retest Reliability of EEG Data

A total of 14 subjects were retested on our EEG measures on two separate occasions. All of these subjects came from the Surrey school, but testing was carried out at the Institute of Psychiatry on the second occasion. Approximately 16 months elapsed between the two testing sessions.

Analysis of Results

The analysis of the data consisted mainly of the computation of simple statistics for all of the variables for each of the main subsamples. The appendices give these results in full. The means and standard deviations reported reflect the output of the computer program used for the analysis, which did not have a facility for specifying an assumed decimal point in the data input records. Analyses are included for the entire school sample, the school men, the school women, the court stenographers, and the Mensa society. In addition to the means, etc., product-moment correlations were computed between all variables. The sample size included in the various statistics varied somewhat because of missing data. The sample sizes for each computed statistic are given in the appendices. The correlation coefficients in the full tables will be seen to *have no underscores, a single underscore, or a double underscore*. These indicate that the correlations were not significantly different from zero, the 1 chance in 20 (.05 level) and the 1 chance in 100 (.01 level) levels of significance respectively.

Table 1 shows the product-moment correlation coefficients between the experimental EEG-based putative intelligence measures and the major WAIS scores. The correlations are taken from the total school sample ($N = 219$ for the majority of the correlations.) The other summary tables in this section are also taken from the school sample unless shown to the contrary.

Examination of Table 1 indicates that our measures had mixed success in terms of their correlations with the WAIS measures. The zigzag score did not correlate very well at all, and henceforth we will not include this score in any of our summary tables in this section. The multiple string measure had quite significant correlations with the WAIS measures. However, as the multiple string measure was surpassed by the other measures, we will again not include the score except in the complete tables. The string measure was highly correlated with the full WAIS IQ (.72) and the variance measure had an almost identical correlation with the full WAIS in the opposite direction. As the string and the variance measures were themselves correlated only to the extent of −.53, they are evidently not exactly the same measure in terms of some underlying entity. The composite score, "variance minus string" proved to have an

Table 1. Product-moment correlation coefficients between the experimental EEG-based putative intelligence measures and the major WAIS scores

Test	String	Variance	Multiple-string	Zigzag	Variance-string	Verbal total	Performance total	WAIS IQ total
String	1.00							
Variance	−.53	1.00						
Multiple	.65	−.46	1.00					
Zigzag	.11	−.04	.04	1.00				
Var-str	−.87	.88	−.63	−.08	1.00			
Verb tot	.68	−.69	.51	.07	−.78	1.00		
Perf tot	.53	−.53	.42	.10	−.60	.54	1.00	
IQ tot	.72	−.72	.57	.09	−.83	.95	.69	1.00

Table 2. Mean scores of EEG-based 'best' measures, the WAIS main scores, and the personality variables for different subsamples

Test	School sample	Mensa sample	Court sample	Schoolboys sample	Schoolgirls sample
No. obs.	218	19	16	121	98
Variance	162	99	115	162	161
String	139	249	197	143	135
Var-string	22.4	−149.8	−81.3	19.2	26.5
Verb tot	107	N/A	128	108	106
Perf tot	107	N/A	121	107	107
IQ tot	108	147	126	108	107
No. obs.	196	10	12	114	83
P	3.94	3.30	2.58	5.03	2.46
E	14.80	11.30	12.33	14.80	14.80
N	10.38	10.00	11.33	8.92	12.37
L	5.42	5.60	8.67	5.14	5.80

even higher level of correlation (−.83) with the full WAIS measure.

Table 2 shows the mean scores of our EEG-based 'best' measures, the WAIS main scores, and the personality variables for our different subsamples. The school sample is also shown divided into boys and girls.

Perhaps the most noteworthy aspect of Table 2 is the scores for the Mensa subsample. The IQ score shown for this subsample is the self-reported Mensa IQ test, and hence is not comparable to the other subsample IQ means. However, the string measure for the Mensa group was very high compared to the school total; 249 v. 139. Likewise, the variance score was significantly lower; 99 vs 162. These results tend to indicate that the Mensa selection test and our EEG based measures may be in close accordance.

The court stenographer sample did well on our measures, and in view of the fact that they had a mean full WAIS IQ of 126, it lends further support to our contention that the EEG-based measures are measuring some aspect of intelligence.

Examination of the full tables of correlations in the appendices shows that the correlation between our composite score and the full WAIS IQ was −.83 for both the total school sample and the court group. Although that exact correspondence is a coincidence, the two tables of correlations agree quite well. That no doubt results from the fact that the court stenographer sample has a good deal of variance on most of the measures included in the study. The correlation between the self-reported IQ measure for the Mensa sample and our composite measure is only .03. Not only is this low, but the correlation is in the opposite direction to what might be expected. This may be due to the reduction in variance in this subsample. However, in view of the fact that the Mensa IQ measure could not be independently verified (except, of course, as a lower bound by virtue of their membership in the society) it is difficult to interpret the correlation.

The two school sex subsamples had almost identical WAIS mean scores. However, the EEG-based measures showed one measure (variance) with no differences, and one (string) with substantial differences between the boys and girls. The composite score was almost necessarily different because of the differences between the string measures. However, examination of the correlations between the EEG measures and the total WAIS IQ within each sex subsample (see appendices) showed almost identical levels of correlations.

The personality measures showed the

Table 3. Relationship between EEG measures and the WAIS subtests

WAIS test	Variance	String	Variance minus string	Full WAIS IQ current study	Full WAIS IQ published data
Information	–.64	.55	–.68	.80	.84
Comp	–.50	.53	–.59	.74	.72
Arith	–.57	.56	–.65	.79	.70
Simil	–.69	.54	–.71	.84	.80
Digit span	–.54	.49	–.59	.71	.61
Vocabulary	–.57	.62	–.68	.79	.83
Verb total	–.69	.68	–.78	.95	.96
Digit sym	–.28	.32	–.35	.45	.68
Pict comp	–.47	.52	–.57	.67	.74
Blocks	–.50	.45	–.54	.70	.72
Pict arr	–.36	.45	–.46	.54	.68
Obj assembly	–.32	.45	–.44	.55	.65
Perf total	–.53	.53	–.60	.69	.93
WAIS total	–.72	.72	–.83	1.00	1.00

Table 4. Comparison of main EEG measures with the full WAIS IQ

Epoch measure	Variance 256 ms	String 256 ms	Variance minus string 256 ms	Variance 512 ms	String 512 ms	Variance minus string 512 ms	WAIS IQ total
Var 256	1.00						
Str 256	–.53	1.00					
V-s 256	.88	–.87	1.00				
Var 512	.60	–.18	.39	1.00			
Str 512	–.21	.67	–.58	–.08	1.00		
V-s 512	.44	–.67	.67	.49	–.91	1.00	
IQ tot	–.72	.72	–.83	–.35	.47	–.56	1.00

boys and girls as being the same in 'E', the girls higher in 'N', and the boys higher in 'P'. Both the court and Mensa samples were somewhat low on the 'E' measure, and the court sample was high on 'L'. As the 'L' items have a face validity of high moral rectitude, we feel the professional cynicism implied in regarding the scale as a 'Lie' scale may be unjustified in view of fact that these people are court officials.

Table 3 shows the relationship between our EEG measures and the WAIS subtests. We also show the relationship between the full WAIS IQ and the subtests as determined by our own data, and the published norms given in the official WAIS manual for the 18- to 19-year age group (Wechsler 1955) for comparison. The Wechsler correlations have been corrected for tautological contamination.

The pattern of correlations shows that our measures correlate rather more highly with the verbal subtests than the performance subtests. However, the same may be said of the full WAIS IQ measure itself. There is quite a close correspondence between the subtest correlations and our composite measure on the one hand, and the subtest correlations and the full WAIS IQ as measured by us, on the other. Our own subtest versus full WAIS correlations agree well with the published ones for the verbal measures, but seem to be rather lower than the published correlations for the performance measures.

Table 4 compares our main EEG mea-

Table 5. Data for estimation of test-retest reliability of EEG measures

Subject code number	Full WAIS IQ	Variance first occasion	Variance second occasion	String first occasion	String second occasion
82A	100	109	104	78	73
173A	102	139	146	145	147
W31	104	112	94	104	110
142A	104	103	103	174	189
111A	105	115	120	182	181
123A	108	116	118	170	213
151A	116	138	137	231	192
183A	117	105	114	113	120
141A	118	112	103	181	82
162A	124	137	144	153	158
122A	124	105	106	162	153
112A	127	114	115	153	160
132A	130	120	143	255	245
91A	131	118	100	156	137

sures, string, variance, and composite, as computed for a 256-ms epoch and a 512-ms epoch, with the full WAIS IQ. These correlations are based on the school sample, but with a reduced N of 78. A subsample was selected rather than the full 219 records available because of the effort involved in rescoring the records.

As can be seen by examination of the bottom row of Table 4, the 256-ms epoch has a substantially higher relationship to full WAIS IQ than the same measures based on a 512-ms epoch. When it is considered that the 512-ms epoch must have a degree of correlation inbuilt because it includes the 256-ms part of the record as well, it may indicate that there is very little relationship, if any, between our measures and WAIS IQ after the first quarter second or so following stimulus onset.

Data were available from only 14 school subjects to estimate test-retest reliability of our EEG measures (Table 5). These are obviously too few to be happy about the quantification of the test-retest reliability. We accordingly show our raw data for the string and variance measures (256-ms epoch) on the two test occasions, which were approximately 16 months apart. Subjects were all drawn from the Surrey school. The data are presented in rank order of the WAIS full scale IQ, which is also shown in the table. The IQ distribution is somewhat abnormal, in that it is skewed towards the upper tail. This would have the effect of attenuating any coefficient computed from the data.

Discussion of Results

In general, we were happy with our main school sample. There is some evidence that we oversampled the higher IQ subjects and oversampled boys. In both cases, this seems to have occurred because of the self-selection aspect involved in our recruitment methods in the schools. This was largely unavoidable. There is also some indication from the personality data that the volunteers tended to be more extraverted and less neurotic than a general student sample. Boys were slightly lower than the Eysenck normative sample on the 'P' dimension, whereas girls were slightly higher than the 'P' norms.

The string measure results obtained in our research are in close agreement with our previous findings. The correlation of .72 for our school sample is not greatly different than the .77 we first computed on the published Ertl and Schafer (1969) data, even though the basis for the correlations were fairly different.

It is also quite interesting to note the differences made by the epoch of the AEP, which are shown in our tables. The string measure correlated .47 with WAIS IQ for the 512-ms epoch for our sample. As it happens, we also used a 512-ms epoch for the study on the 37 psychology students mentioned in the introduction. It may be that the very close correspondence of the .47 correlation computed by Blinkhorn between our string measure and the RAM IQ scores of his sample is a function of the same epoch being used in the two samples. (The Ertl and Schafer epoch was 250 ms, which tends to substantiate our belief.)

Blinkhorn noted that the reduction in variance of the RAM scores might have attenuated the reported correlation between the string measure and the RAM IQ scores, but perhaps it is more likely that the use of the 512-ms epoch had more to do with the reduced correlation. Both factors must be taken into account when interpreting these results.

Although there is not a lot of published literature comparing AEP epochs, there is one study by Osborne on the reliability of the AEP which is very relevant to this question (Osborne 1970). Using a sample comprising 13 pairs of Mz twins and six pairs of Dz twins, Osborne compared the waveforms of visual evoked responses taken from the same people on two occasions some 17 weeks apart. The waveforms were compared by correlating the corresponding A/D points. The waveforms were divided into three epochs of 250 ms each. As Osborne reports, 'The median *r* for the first 250 ms was .94 for the middle third of the tracing, .62, for the last third of the plot the median *r* was $-.07$. From these results it is clear that the visual evoked response is stable over Time but not all parts of the tracing are equally congruent'.

All in all, there is some evidence that the AEP waveform represents the activity caused by the initial pulse trains initiated by the receptor cells as they are propogated throughout the cortex.

The string and the variance measures seem to have approximately the same degree of relationship to the WAIS measures. The fact that the variance measure does not show any differences in the mean scores of our sex subsamples may indicate that it is to be preferred to the string measure.

The theory in the preceding chapter has stated that the AEP waveform is a function of the pulse trains initiated by the stimulus, but replicated throughout large numbers of nerve fibres. The larger brain size of men might therefore account for waveforms with larger amplitudes, if we assumed that some of the 'extra' material represented larger populations of such 'primary' fibres. Larger waveform amplitudes, all other things being equal, would result in string scores of a greater magnitude.

The standard deviations of the string measure were nearly the same for the two sexes, although the mean score of the boys was higher. The opposite pattern occurs for the variance measure, where the mean scores are very close, but the standard deviations are different.

The variance measure shows a standard deviation of 49.869 for the girls and 59.037 for the boys. As was noted in the previous chapter, a number of IQ tests have shown boys to have larger standard deviations than girls. Although the ratio 49.869/59.037 may not seem to be very significant at first sight, it has a large effect on the absolute numbers of women and men that will be found in the tails of IQ distribution. Table 5 shows some calculations we carried out on the basis of the above-mentioned differences in the 'variance' standard deviations, showing the expected numbers of each sex in the United Kingdom that would be found to have scores exceeding the one to five multiples of the combined sexes standard deviation. (The U.K. population is taken as a round 55,000,000.)

Table 6 shows that the ratio of men to women changes markedly as one goes up the IQ scale. At IQ 130, the level thought

Table 6. Expected numbers of each sex in the United Kingdom with scores exceeding the one to five multiples of the combined sexes standard deviation

Standard deviation from mean[a]	Male variance ratio	Expected number of men in pop. of 27500000	Female variance ratio	Expected number of women in pop. of 27500000
(115) 1	.9317	4833003	1.1029	3713460
(130) 2	1.8633	858288	2.2059	376680
(145) 3	2.7950	71463	3.3088	12960
(160) 4	3.7266	2709	4.4117	149
(175) 5	4.6582	47	5.5146	1

[a] The equivalent WAIS IQ scores are shown in brackets (the WAIS IQ is deliberately scaled to have a standard deviation of 15).

to represent the approximate mean of students undertaking postgraduate studies at university, men outnumber women by more than 2 to 1. At IQ 145 and above, the ratio becomes 5.5 men for every woman. At IQ 160 and above, there are 18 men for every woman. Finally, we would expect to find only one woman in 27,500,000 with an IQ of 175 or more whereas there would be 47 men.

If the variance score standard deviations were to be regarded as some indication of the true differences in the distribution of intelligence in the two sexes, it would probably not be necessary to look for further reasons to explain the relative lack of women in higher occupational levels.

The 'multiple string' measure had quite respectable correlations with the WAIS measures, and perhaps it is only because the variance, string, and composite measures are better correlated that we have not paid more attention to it. Why the multiple string measure should work at all has not been discussed in detail. However, we believe that it results from the fact that even a single EEG sample is in fact an 'average', if one accepts our interpretation that the electrogenesis of the EEG is a summation of the firing of the thousands of time congruent fibres. This raises the interesting possibility that it may not be necessary to have an 'evoked response' paradigm in order to obtain an EEG-based measure of intelligence. It may be that all one has to do is to seat the subject in a 'consistent stimulus' environment (say, very quiet, or perhaps with a white noise presented continuously) and record background EEG for a fairly long epoch. The continuous string measure could then be computed from the entire EEG record. In view of our results, we feel it is likely that this measure would correlate well with standard IQ measures.

The composite measure proved to have the highest correlation (−.83) with the full WAIS IQ measure. If it were not for the fact that the means of the two sex subsamples differed on the string measure and the composite score, we would recommend the adoption of that score as being the best available EEG-based measure of IQ. At the moment, taking all of our results into consideration, we feel that the variance measure is the best single EEG-based measure of intelligence.

There is not much to say about the WAIS subtest correlations with our measures beyond noting again that the correlation pattern was not too dissimilar to the overall WAIS IQ itself. The largest differences were in the area of the performance measures, which correlate far less well with our measures than do the verbal subtests. Although we believe that our EEG-based measures are measuring 'fluid' intelligence, the previous chapter has pointed out that fluid and crystallized intelligence themselves would be expected to be highly correlated within any given homogeneous environment. We have no reason for supposing that the WAIS performance measures are in any sense better measures of fluid intelligence than the verbal measures.

The high mean scores of the EEG-related measures that were obtained for our two special subsamples give us additional confidence in the construct validity of our EEG measures over and above the high levels of correlations that we obtained.

No results have been presented for the 'severely subnormal' (SSN) group of 15 subjects that were tested. Partly, this is due to the fact that testing this group proved to be difficult. The subjects were very cooperative as far as their affect was concerned, but it was very difficult to get them to sit still during the recording sessions and to prevent them from removing their electrodes due to simple curiosity. Nonetheless, a number of reasonable recordings were obtained from this sample and processed by us to the extent of looking at the individual records and computing some of our scores.

We found that the individual SSN EEG records have a very different appearance to any of our normal subsamples. The spiky nature of the individual records tends to disappear, and what one sees instead are smoother but quite large changes in the EEG signal.

Although we can only speculate, we feel that what we are seeing with the SSN group are records which do not show a consistency of pulse train firing, because of the smoothness of the waveform over the epoch. This, of course, would be consistent with our other interpretations. However, how can we account for the very large amplitudes of the EEG signal, even if the spiky nature is missing? This, we feel, may be because there is very little else going on in the cortex (relative to our other subjects, of course) apart from the transmission of the primary stimulus. If the level of recognition probability 'R' is greatly reduced for these people, it may be that very few associated pulse trains are ever initiated. Hence, the majority of the EEG activity would be a reflection of the simple transmission of the primary stimulus inputs.

If the above reasoning is correct, it would mean that the lack of other neurons firing might result in the high amplitude signal we do measure. The EEG amplifier amplifies the difference in voltage between two points, and if there is a lot of unrelated electrical activity at the two electrodes, it should tend to cancel out overall, but 'damp' the other activity at the same time. The damping effect of this associated neural firing may be missing in this special group.

In any event, the abnormal records, when averaged and run through our programs, tended to produce very large string measures; larger, in some cases, than our Mensa subjects.

This processing was not carried out on any data that would allow us to make actual comparisons of a numeric sort between the SSN data and any of our other data. This is because of the fact that we could not apply our consistent editing procedures to this data. Many of the SSN EEG records had to be rejected because of amplifier limiting, which is easily seen as a constant (flat line) voltage output from the amplifier.

Because of the highly inconsistent (and low) numbers of acceptable records we had for this sample, we did not compute any of the other EEG measures.

In summary, we can say that the EEG seems able to detect this sort of mental abnormality, but the string measure in particular does not have any construct validity for this end of the IQ spectrum. Any use of our measures, therefore, should be applied with some screening of the sample if there is any prior possibility that subjects such as these might be included in a larger sample. It may be the case, however, that had we adjusted our amplifier gain appropriately prior to recording and scaled the results after, that the variance measure would have shown this group to have the highest variance scores of all (which indicate low IQ). Further data will have to be collected to establish this for certain.

A fair number of studies have appeared in the literature which have not shown any relationship between EEG-based measures and conventional IQ tests. As we are personally familiar with the details of the methodology employed by some of these failures, we are not too surprised that the failures should have occurred. Our own failure with the current data when we introduced our abortive 'automated editing procedure' has shown us just how careful one has to be to consider the possible effect of even a small change in procedure on the eventual results.

As our discourse about methodology and processing is scattered throughout two chapters, we feel it may be of benefit to present a brief summary of what should be done by anybody wishing to duplicate our results.

1. Stimulus Choice. It is very important to use a stimulus that is completely constant from one presentation to the next, considered at the level of individual receptor cells that will be firing. Ideally, the same population of receptor cells should fire with the same pulse train time pattern with each presentation of the stimulus. We avoided visual stimuli for that reason, because of the virtual impossibility of controlling eye fixation to the necessary degree of accuracy. Likewise, we would not expect an auditory 'click' stimulus to work, as the frequency content of a click is not consistent from one presentation to the next. Perhaps the worst possible stimulus for our special purposes would be the visual 'reverse checkerboard' that is used in some AEP research.

If an auditory stimulus is used, care should be taken to switch it in and out of circuit at zero crossings, to avoid the production of a 'click' that might otherwise occur.

2. Stimulus presentation. The interstimulus interval should be varied on a pseudorandom basis from one presentation to the next, to prevent habituation effects. However, the same random sequence should be presented to each subject.

3. Electrodes. We feel happiest with the silver/silver chloride electrodes. Attachment must be very secure, and scalp abrasion done with great care.

4. Amplifier. We used our own special amplifier, without any special upper frequency filtering, as we wished to record all natural fast occurrences of voltage change. Our interpretation of the EEG as reflecting pulse train activity means that we should be able to detect the rise and fall of individual spikes if possible. The cost of our special amplifier was less than 10 pounds, and it was constructed in under 4 h. The circuit diagram is shown in Appendix N.

5. Calibration signal. A constant calibration voltage must be fed into the record for each subject (or provided to an on-line program if processed in real time.) The final conversion of the signals to digital values must ensure that the calibration signal is represented as a constant sum for each subject. Failure to do this will apply an unknown scaling factor to the scores of each subject, and completely destroy their validity.

6. Recording medium. Ideally, signals should be directly converted as they are obtained and stored in digital form at the onset. As that may be difficult in remote testing sites, recording of the data must be done on recorders with a very high specification of speed stability. We were interested to notice that we obtained slightly inconsistent results from rescoring the same data from our subjects on different occasions. This proved to be caused by variations in tape speed. It was also of interest to note that the differences became larger as one went further into the record, in accordance with the simulation study mentioned in the previous chapter.

7. A/D conversion sampling rate. Again, we think that this should be fast enough to pick up any genuine activity recorded at the electrodes. We feel that 1 ms should be regarded as a minimum period, rather than a recommended one. Our 512 ms epoch study utilized a 2-ms sampling period, and ideally, we should not have done this.

8. *Epoch of analysis period.* This should be 250 ms for all of the reasons stated previously. Longer epochs will invalidate the assumptions underlying our measures. It probably would not be harmful to have shorter epochs, but this may be a waste of data.

9. *Editing of individual records.* If there is a lot of 50 (60 U.S.A.) cycle interference on the records, the entire session should be scrapped. If possible, an inspection of the records should be made at the time they are obtained, to avoid excessive data loss.

On no account should 'somewhat abnormal' records be rejected just because of that fact. To do so might be to build in a consistent bias in terms of our scores, as we ourselves found in this research.

The final number of records retained for the purposes of averaging must be constant for each subject. All other things being equal, a smaller number of records will tend to produce larger string scores, and the measure is very sensitive to this.

Our editing procedure of rejecting the first few records in each session (those tending to show muscle artefact) is recommended. Records at the end of the testing session should be rejected to keep the number of records used in computing the EEG measures exactly the same for each subject.

Acknowledgments. The work of the author has been supported by grant HR4925/I from the Social Science Research Council. We would like to thank the Council for their support. We would also like to thank Peter Smith of PTT Ltd., who designed all of the special equipment used in this study.

Appendix A. Summary statistics and correlations of the school sample: summary statistics

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Variance-256	161.776	55.002	3025.2	3.7167	219.
String-256	139.338	53.827	2897.3	3.6373	219.
Multiple	156.260	56.650	3209.2	3.8281	219.
Z score	33.836	3.188	10.2	0.2154	219.
Sex	1.563	0.497	0.2	0.0339	215.
Age	15.670	1.135	1.3	0.0774	215.
P	3.944	2.909	8.5	0.2072	197.
E	14.797	4.099	16.8	0.2921	197.
N	10.376	4.352	18.9	0.3101	197.
L	5.416	3.346	11.2	0.2384	197.
Variance-512	145.782	29.499	870.2	3.3401	78.
String-512	210.590	60.908	3709.8	6.8965	78.
Information	9.679	3.158	10.0	0.2139	218.
Comp	11.518	3.504	12.3	0.2373	218.
Arith	9.372	2.955	8.7	0.2001	218.
Simil	11.399	3.210	10.3	0.2174	218.
Digit span	10.128	3.548	12.6	0.2403	218.
Vocabulary	10.500	3.066	9.4	0.2077	218.
Verb tot	107.298	15.464	239.1	1.0473	218.
Digit symb	10.945	2.710	7.3	0.1835	218.
Pict comp	11.261	2.468	6.1	0.1672	218.
Blocks	11.220	2.875	8.3	0.1947	218.
Pict arr	10.326	2.326	5.4	0.1575	218.
Obj assemb	10.225	3.172	10.1	0.2149	218.
Perf tot	107.092	15.176	230.3	1.0278	218.
IQ tot	107.662	13.910	193.5	0.9400	219.
Var-str 256	22.438	95.141	9051.9	6.4291	219.
Var-str 512	−64.808	69.673	4854.3	7.8889	78.

Appendix B. Summary statistics and correlations of the school sample: product-moment correlation

	1	2	3	4	5	6	7	8	9	10	11	12
1 Variance-256	1.00											
2 String-256	**−0.53**	1.00										
3 Multiple	**−0.46**	**0.65**	1.00									
4 Z score	−0.04	0.11	0.04	1.00								
5 Sex	−0.01	0.10	0.05	−0.02	1.00							
6 Age	−0.03	0.07	*0.12*	*−0.14*	0.06	1.00						
7 P	0.02	*−0.13*	−0.07	0.04	**0.44**	−0.04	1.00					
8 E	0.03	−0.03	−0.05	−0.01	0.00	**−0.17**	0.11	1.00				
9 N	−0.02	−0.01	0.00	**0.18**	**−0.39**	−0.05	*−0.15*	**−0.19**	1.00			
10 L	0.08	−0.05	−0.03	−0.01	−0.10	−0.03	**−0.29**	*−0.12*	−0.08	1.00		
11 Variance-512	**0.60**	−0.18	**−0.39**	−0.03	0.04	−0.18	0.09	0.07	0.17	*−0.24*	1.00	
12 String-512	*−0.21*	**0.67**	**0.48**	0.00	0.04	0.17	*−0.21*	0.03	−0.11	0.00	−0.08	1.00
13 Information	**−0.64**	**0.55**	**0.43**	0.09	**0.17**	0.03	0.07	−0.01	−0.11	−0.11	**−0.27**	0.17
14 Comp	**−0.50**	**0.53**	**0.42**	0.03	−0.01	0.02	−0.05	−0.02	0.04	−0.08	**−0.34**	0.17
15 Arith	**−0.57**	**0.56**	**0.42**	−0.02	*0.13*	−0.03	0.04	0.06	−0.10	−0.02	**−0.28**	*0.26*
16 Simil	**−0.69**	**0.54**	**0.36**	**0.18**	0.03	−0.07	0.07	0.11	−0.00	−0.11	*−0.26*	*0.26*
17 Digit span	**−0.54**	**0.49**	**0.39**	−0.03	*0.12*	−0.04	0.03	0.03	**−0.18**	−0.07	*−0.24*	**0.37**
18 Vocabulary	**−0.57**	**0.62**	**0.50**	*0.13*	0.08	0.03	0.06	−0.03	*−0.12*	−0.08	**−0.28**	*0.22*
19 Verb tot	**−0.69**	**0.68**	**0.51**	0.07	0.09	−0.04	0.04	0.04	−0.09	−0.10	**−0.35**	**0.31**
20 Digit symb	**−0.28**	**0.32**	**0.20**	−0.03	−0.07	−0.03	−0.04	0.04	0.11	−0.06	−0.12	0.16
21 Pict comp	**−0.47**	**0.52**	**0.40**	0.08	0.03	0.08	−0.07	−0.07	−0.01	−0.09	**−0.39**	**0.39**
22 Blocks	**−0.50**	**0.45**	**0.40**	0.09	0.09	0.09	−0.08	*0.16*	−0.01	−0.05	−0.17	**0.27**
23 Pict arr	**−0.36**	**0.45**	**0.32**	*0.12*	0.03	0.07	*−0.16*	−0.07	0.07	−0.09	−0.01	**0.36**
24 Obj assemb	**−0.32**	**0.45**	**0.42**	*0.15*	0.04	0.10	*−0.14*	−0.11	−0.03	0.01	−0.18	**0.29**
25 Perf tot	**−0.53**	**0.53**	**0.42**	0.10	0.01	0.01	−0.07	−0.04	0.09	−0.01	**−0.28**	**0.47**
26 IQ tot	**−0.72**	**0.72**	**0.57**	0.09	0.08	0.01	−0.04	0.01	−0.04	−0.10	**−0.35**	**0.47**
27 Var-str 256	**0.88**	**−0.87**	**−0.63**	−0.08	−0.06	−0.06	0.09	0.04	−0.01	0.08	**0.39**	**−0.58**
28 Var-str 512	**0.44**	**−0.67**	**−0.58**	−0.01	−0.02	*−0.23*	*0.22*	0.00	0.17	−0.11	**0.49**	**−0.91**

Appendix C. Summary statistics and correlations of the school sample: sample base of correlation

	1	2	3	4	5	6	7	8	9	10	11	12
5 Sex	215.	215.	215.	215.	215.							
6 Age	215.	215.	215.	215.	215.	215.						
7 P	197.	197.	197.	197.	197.	197.	197.					
8 E	197.	197.	197.	197.	197.	197.	197.	197.				
9 N	197.	197.	197.	197.	197.	197.	197.	197.	197.			
10 L	197.	197.	197.	197.	197.	197.	197.	197.	197.	197.		
11 Variance-512	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.	
12 String-512	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.
13 Information	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
14 Comp	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
15 Arith	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
16 Simil	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
17 Digit span	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
18 Vocabulary	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
19 Verb tot	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
20 Digit symb	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
21 Pict comp	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
22 Blocks	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
23 Pict arr	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
24 Obj assemb	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
25 Perf tot	218.	218.	218.	218.	214.	214.	196.	196.	196.	196.	77.	77.
26 IQ tot	219.	219.	219.	219.	215.	215.	197.	197.	197.	197.	78.	78.
27 Var-str 256	219.	219.	219.	219.	215.	215.	197.	197.	197.	197.	78.	78.
28 Var-str 512	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.	78.

coefficients

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1.00															
0.55	1.00														
0.66	0.51	1.00													
0.69	0.63	0.63	1.00												
0.55	0.41	0.58	0.61	1.00											
0.70	0.68	0.61	0.69	0.62	1.00										
0.82	0.77	0.80	0.85	0.77	0.87	1.00									
0.20	0.40	0.31	0.31	0.18	0.21	0.33	1.00								
0.50	0.49	0.50	0.50	0.46	0.47	0.59	0.26	1.00							
0.51	0.37	0.52	0.58	0.38	0.38	0.55	0.35	0.41	1.00						
0.36	0.29	0.38	0.36	0.23	0.40	0.40	0.31	0.28	0.39	1.00					
0.34	0.28	0.32	0.35	0.24	0.31	0.37	0.27	0.39	0.51	0.41	1.00				
0.45	0.44	0.51	0.51	0.35	0.43	0.54	0.45	0.50	0.66	0.53	0.59	1.00			
0.80	0.74	0.79	0.84	0.71	0.79	0.95	0.45	0.67	0.70	0.54	0.55	0.69	1.00		
−0.68	−0.59	−0.65	−0.71	−0.59	−0.68	−0.78	−0.35	−0.57	−0.54	−0.46	−0.44	−0.60	−0.83	1.00	
−0.26	−0.29	−0.34	−0.34	−0.43	−0.31	−0.42	−0.19	−0.51	−0.31	−0.32	−0.33	−0.53	−0.56	0.67	1.00

coefficients[a]

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
218.															
218.	218.														
218.	218.	218.													
218.	218.	218.	218.												
218.	218.	218.	218.	218.											
218.	218.	218.	218.	218.	218.										
218.	218.	218.	218.	218.	218.	218.									
218.	218.	218.	218.	218.	218.	218.	218.								
218.	218.	218.	218.	218.	218.	218.	218.	218.							
218.	218.	218.	218.	218.	218.	218.	218.	218.	218.						
218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.					
218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.				
218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.			
218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	219.		
218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	218.	219.	219.	
77.	77.	77.	77.	77.	77.	77.	77.	77.	77.	77.	77.	77.	78.	78.	78.

[a] Unless shown to the contrary, correlations are based on a sample size of 219.

Appendix D. Summary statistics and correlations for the schoolgirls sample: summary statistics

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Variance-256	161.347	49.869	2486.9	5.0375	98.
String-256	134.878	53.601	2873.0	5.4145	98.
Multiple	154.112	57.755	3335.6	5.8341	98.
Z score	33.878	3.091	9.6	0.3122	98.
Sex	1.000	0.000	0.0	0.0000	94.
Age	15.596	0.738	0.5	0.0761	94.
P	2.458	2.050	4.2	0.2250	83.
E	14.795	4.364	19.0	0.4790	83.
N	12.373	3.747	14.0	0.4113	83.
L	5.795	3.327	11.1	0.3652	83.
Variance-512	144.500	30.522	931.6	5.3956	32.
String-512	207.844	59.361	3523.7	10.4936	32.
Information	9.153	3.030	9.2	0.3061	98.
Comp	11.663	3.652	13.3	0.3689	98.

Appendix E. Summary statistics and correlations for the schoolgirls sample: product-moment correla-

	1	2	3	4	5	6	7	8	9	10	11	12
1 Variance-256	1.00											
2 String-256	**–0.50**	1.00										
3 Multiple	**–0.48**	**0.70**	1.00									
4 Z score	–0.13	*0.21*	0.11	1.00								
5 Sex	0.00	0.00	0.00	0.00	1.00							
6 Age	0.15	–0.01	0.03	–0.14	0.00	1.00						
7 P	*0.19*	*–0.22*	–0.14	–0.18	0.00	–0.09	1.00					
8 E	0.01	0.10	0.04	0.09	0.00	–0.13	0.17	1.00				
9 N	*–0.25*	0.15	*0.22*	0.08	0.00	*–0.19*	0.16	–0.12	1.00			
10 L	0.05	–0.08	–0.15	0.07	0.00	0.16	*–0.24*	–0.10	–0.06	1.00		
11 Variance-512	**0.70**	–0.24	*–0.37*	–0.02	0.00	–0.11	0.19	–0.01	0.14	*–0.38*	1.00	
12 String-512	–0.17	**0.70**	**0.56**	0.19	0.00	0.09	*–0.30*	0.13	0.03	0.07	0.07	1.00
13 Information	**–0.55**	**0.49**	**0.42**	0.09	0.00	0.00	–0.08	–0.07	0.13	*–0.20*	–0.18	0.11
14 Comp	**–0.54**	**0.58**	**0.48**	0.02	0.00	*–0.23*	–0.02	0.12	0.09	–0.13	–0.28	0.25
15 Arith	**–0.56**	**0.55**	**0.44**	0.12	0.00	–0.04	–0.01	0.15	0.18	*–0.19*	–0.26	0.09
16 Simil	**–0.69**	**0.58**	**0.42**	**0.25**	0.00	*–0.19*	–0.00	0.18	*0.21*	–0.09	–0.24	0.29
17 Digit span	**–0.55**	**0.57**	**0.46**	0.10	0.00	–0.01	–0.10	0.09	0.11	0.01	*–0.31*	**0.41**
18 Vocabulary	**–0.58**	**0.59**	**0.45**	0.15	0.00	0.01	0.01	0.09	0.02	–0.10	*–0.35*	0.13
19 Verb tot	**–0.68**	**0.69**	**0.55**	0.14	0.00	–0.11	–0.03	0.11	0.16	–0.14	*–0.33*	0.29
20 Digit symb	**–0.33**	**0.39**	**0.32**	0.06	0.00	**–0.24**	0.16	*0.23*	*0.21*	**–0.34**	–0.17	*0.35*
21 Pict comp	**–0.43**	**0.51**	**0.45**	*0.22*	0.00	–0.07	–0.09	0.04	*0.18*	*–0.22*	*–0.32*	0.19
22 Blocks	**–0.46**	**0.44**	**0.46**	**0.24**	0.00	**–0.24**	0.02	0.17	*0.20*	–0.17	–0.16	*0.40*
23 Pict arr	**–0.41**	**0.46**	**0.40**	*0.23*	0.00	–0.06	**–0.29**	0.10	*0.24*	–0.16	0.05	**0.45**
24 Obj assemb	**–0.34**	**0.47**	**0.39**	*0.19*	0.00	–0.10	*–0.23*	–0.07	0.10	–0.05	*–0.34*	**0.53**
25 Perf tot	**–0.51**	**0.45**	**0.37**	*0.18*	0.00	**–0.30**	0.01	0.03	**0.33**	–0.09	–0.29	**0.56**
26 IQ tot	–0.71	**0.72**	**0.60**	*0.21*	0.00	–0.16	–0.09	0.13	*0.22*	*–0.19*	*–0.35*	**0.47**
27 Var-str 256	**0.86**	**–0.88**	**–0.68**	*–0.20*	0.00	0.09	*0.24*	–0.06	*–0.23*	0.07	**0.47**	**–0.59**
28 Var-str 512	**0.48**	**–0.75**	**–0.69**	–0.18	0.00	–0.13	*0.37*	–0.12	0.04	–0.24	*0.41*	**–0.88**

Appendix D (continued)

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Arith	9.031	2.775	7.7	0.2803	98.
Simil	11.378	3.038	9.2	0.3068	98.
Digit span	9.755	3.262	10.6	0.3295	98.
Vocabulary	10.316	2.775	7.7	0.2804	98.
Verb tot	106.173	14.804	219.2	1.4954	98.
Digit symb	11.214	2.923	8.5	0.2952	98.
Pict comp	11.204	2.311	5.3	0.2334	98.
Blocks	11.000	2.663	7.1	0.2690	98.
Pict arr	10.296	2.557	6.5	0.2583	98.
Obj assemb	10.143	2.872	8.2	0.2901	98.
Perf tot	107.153	17.750	315.1	1.7931	98.
IQ tot	106.796	13.422	180.1	1.3558	98.
Var-str 256	26.469	89.712	8048.3	9.0623	98.
Var-str 512	–63.344	64.773	4195.5	11.4503	32.

tion coefficients

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1.00															
0.52	1.00														
0.61	**0.65**	1.00													
0.59	**0.67**	**0.71**	1.00												
0.57	**0.51**	**0.58**	**0.63**	1.00											
0.68	**0.71**	**0.69**	**0.74**	**0.61**	1.00										
0.77	**0.83**	**0.84**	**0.86**	**0.77**	**0.88**	1.00									
0.25	**0.40**	**0.34**	**0.41**	**0.28**	**0.31**	**0.41**	1.00								
0.43	**0.53**	**0.57**	**0.48**	**0.36**	**0.45**	**0.57**	*0.23*	1.00							
0.37	**0.48**	**0.44**	**0.65**	**0.44**	**0.44**	**0.55**	**0.44**	**0.42**	1.00						
0.37	**0.39**	**0.40**	**0.49**	**0.26**	**0.38**	**0.46**	**0.39**	**0.31**	**0.46**	1.00					
0.24	**0.28**	**0.25**	**0.37**	**0.27**	*0.21*	**0.34**	**0.31**	**0.39**	**0.51**	**0.47**	1.00				
0.30	**0.43**	**0.41**	**0.48**	**0.27**	**0.37**	**0.47**	**0.41**	**0.36**	**0.61**	**0.48**	**0.45**	1.00			
0.73	**0.79**	**0.80**	**0.87**	**0.72**	**0.80**	**0.95**	**0.53**	**0.65**	**0.71**	**0.61**	**0.53**	**0.58**	1.00		
–0.60	**–0.65**	**–0.64**	**–0.73**	**–0.65**	**–0.68**	**–0.79**	**–0.41**	**–0.54**	**–0.52**	**–0.50**	**–0.47**	**–0.55**	**–0.82**	1.00	
–0.18	*–0.36*	–0.21	*–0.38*	**–0.52**	–0.28	**–0.42**	*–0.41*	*–0.33*	**–0.44**	*–0.38*	**–0.64**	**–0.65**	**–0.59**	**0.76**	1.00

Appendix F. Summary statistics and correlations for the schoolgirls sample: sample base of correlation

	1	2	3	4	5	6	7	8	9	10	11	12
5 Sex	94.	94.	94.	94.	94.							
6 Age	94.	94.	94.	94.	94.	94.						
7 P	83.	83.	83.	83.	83.	83.	83.					
8 E	83.	83.	83.	83.	83.	83.	83.	83.				
9 N	83.	83.	83.	83.	83.	83.	83.	83.	83.			
10 L	83.	83.	83.	83.	83.	83.	83.	83.	83.	83.		
11 Variance-512	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	
12 String-512	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.
13 Information	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
14 Comp	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
15 Arith	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
16 Simil	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
17 Digit span	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
18 Vocabulary	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
19 Verb tot	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
20 Digit symb	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
21 Pict comp	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
22 Blocks	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
23 Pict arr	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
24 Obj assemb	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
25 Perf tot	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
26 IQ tot	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
27 Var-str 256	98.	98.	98.	98.	94.	94.	83.	83.	83.	83.	32.	32.
28 Var-str 512	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.

[a] Unless shown to the contrary, correlations are based on a sample size of 98.

Appendix G. Summary statistics and correlations for the schoolboys sample: summary statistics

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Variance-256	162.124	59.037	3485.3	5.3670	121.
String-256	142.950	53.960	2911.7	4.9055	121.
Multiple	158.000	55.920	3127.0	5.0836	121.
Z score	33.802	3.278	10.7	0.2980	121.
Sex	2.000	0.000	0.0	0.0000	121.
Age	15.727	1.366	1.9	0.1242	121.
P	5.026	2.970	8.8	0.2782	114.
E	14.798	3.915	15.3	0.3667	114.
N	8.921	4.195	17.6	0.3929	114.
L	5.140	3.347	11.2	0.3135	114.
Variance-512	146.674	29.073	845.2	4.2866	46.
String-512	212.500	62.541	3911.4	9.2212	46.
Information	10.108	3.207	10.3	0.2927	120.
Comp	11.400	3.389	11.5	0.3094	120.

coefficients[a]

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
98.															
98.	98.														
98.	98.	98.													
98.	98.	98.	98.												
98.	98.	98.	98.	98.											
98.	98.	98.	98.	98.	98.										
98.	98.	98.	98.	98.	98.	98.									
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
98.	98.	98.	98.	98.	98.	98.	98.								
32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.	32.

Appendix G. (continued)

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Arith	9.650	3.078	9.5	0.2810	120.
Simil	11.417	3.357	11.3	0.3065	120.
Digit span	10.433	3.752	14.1	0.3425	120.
Vocabulary	10.650	3.289	10.8	0.3002	120.
Verb tot	108.217	15.985	255.5	1.4593	120.
Digit symb	10.725	2.514	6.3	0.2295	120.
Pict comp	11.308	2.599	6.8	0.2372	120.
Blocks	11.400	3.036	9.2	0.2771	120.
Pict arr	10.350	2.129	4.5	0.1943	120.
Obj assemb	10.292	3.409	11.6	0.3112	120.
Perf tot	107.042	12.772	163.1	1.1659	120.
IQ tot	108.364	14.310	204.8	1.3009	121.
Var-str 256	19.174	99.572	9914.5	9.0520	121.
Var-str 512	–65.826	73.576	5413.4	10.8482	46.

Appendix H. Summary statistics and correlations for the school boys sample: product-moment correla-

	1	2	3	4	5	6	7	8	9	10	11	12
1 Variance-256	1.00											
2 String-256	**–0.55**	1.00										
3 Multiple	**–0.45**	**0.62**	1.00									
4 Z score	0.02	0.04	–0.02	1.00								
5 Sex	0.00	0.00	0.00	0.00	1.00							
6 Age	–0.10	0.09	*0.16*	–0.14	0.00	1.00						
7 P	–0.05	*–0.16*	–0.08	*0.16*	0.00	–0.07	1.00					
8 E	0.05	–0.14	–0.13	–0.07	0.00	*–0.19*	0.10	1.00				
9 N	0.10	–0.07	–0.11	**0.23**	0.00	0.03	–0.03	**–0.28**	1.00			
10 L	0.10	–0.03	0.08	–0.06	0.00	–0.09	**–0.31**	–0.14	*–0.18*	1.00		
11 Variance-512	**0.52**	–0.14	**–0.41**	–0.03	0.00	–0.23	0.04	0.13	*0.25*	–0.12	1.00	
12 String-512	–0.24	**0.66**	**0.43**	–0.10	0.00	0.20	*–0.28*	–0.04	–0.16	–0.04	–0.18	1.00
13 Information	**–0.71**	**0.59**	**0.44**	0.09	0.00	0.03	0.05	0.04	*–0.17*	–0.03	**–0.38**	0.22
14 Comp	**–0.48**	**0.50**	**0.37**	0.04	0.00	0.14	–0.05	–0.14	–0.02	–0.05	**–0.37**	0.12
15 Arith	**–0.58**	**0.57**	**0.40**	–0.11	0.00	–0.04	–0.02	–0.00	*–0.19*	0.12	*–0.31*	**0.38**
16 Simil	**–0.69**	**0.52**	**0.32**	0.14	0.00	–0.03	0.10	0.06	–0.11	–0.12	*–0.29*	0.24
17 Digit span	**–0.55**	**0.44**	**0.34**	–0.11	0.00	–0.07	0.03	–0.02	**–0.30**	–0.10	–0.23	**0.36**
18 Vocabulary	**–0.57**	**0.65**	**0.54**	0.11	0.00	0.03	0.05	–0.11	*–0.17*	–0.05	*–0.26*	*0.26*
19 Verb tot	**–0.70**	**0.66**	**0.48**	0.01	0.00	–0.03	0.02	–0.02	*–0.19*	–0.06	**–0.38**	*0.32*
20 Digit symb	**–0.25**	**0.28**	0.09	–0.11	0.00	0.07	–0.09	–0.14	–0.01	*0.16*	–0.06	0.05
21 Pict comp	**–0.50**	**0.53**	**0.37**	–0.01	0.00	0.14	–0.10	*–0.16*	–0.12	0.00	**–0.43**	**0.51**
22 Blocks	**–0.52**	**0.46**	**0.36**	–0.01	0.00	*0.20*	*–0.22*	*0.16*	–0.06	0.03	–0.19	0.18
23 Pict arr	**–0.33**	**0.44**	**0.24**	0.02	0.00	0.15	–0.14	**–0.23**	–0.03	–0.02	–0.05	*0.32*
24 Obj assemb	**–0.30**	**0.43**	**0.45**	0.12	0.00	*0.17*	–0.15	–0.14	–0.08	0.05	–0.08	0.15
25 Perf tot	**–0.58**	**0.62**	**0.48**	0.02	0.00	*0.20*	*–0.18*	–0.13	–0.12	0.08	*–0.28*	**0.42**
26 IQ tot	**–0.73**	**0.73**	**0.54**	0.00	0.00	0.07	–0.08	–0.08	–0.14	–0.04	**–0.36**	**0.46**
27 Var-str 256	**0.89**	**–0.87**	**–0.60**	–0.01	0.00	–0.11	0.06	0.11	0.09	0.07	*0.32*	**–0.57**
28 Var-str 512	**0.41**	**–0.61**	**–0.53**	0.07	0.00	*–0.26*	*0.25*	0.09	0.24	–0.02	**0.55**	**–0.92**

Appendix I. Summary statistics and correlations for the schoolboys sample: sample base of correlation

	1	2	3	4	5	6	7	8	9	10	11	12
7 P	114.	114.	114.	114.	114.	114.	114.					
8 E	114.	114.	114.	114.	114.	114.	114.	114.				
9 N	114.	114.	114.	114.	114.	114.	114.	114.	114.			
10 L	114.	114.	114.	114.	114.	114.	114.	114.	114.	114.		
11 Variance-512	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.	
12 String-512	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.
13 Information	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
14 Comp	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
15 Arith	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
16 Simil	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
17 Digit span	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
18 Vocabulary	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
19 Verb tot	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
20 Digit symb	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
21 Pict comp	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
22 Blocks	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
23 Pict arr	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
24 Obj assemb	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
25 Perf tot	120.	120.	120.	120.	120.	120.	113.	113.	113.	113.	45.	45.
26 IQ tot	121.	121.	121.	121.	121.	121.	114.	114.	114.	114.	46.	46.
27 Var-Str 256	121.	121.	121.	121.	121.	121.	114.	114.	114.	114.	46.	46.
28 Var-str 512	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.	46.

[a] Unless shown to the contrary, correlations are based on a sample size of 121.

tion coefficients

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1.00															
0.59	1.00														
0.69	**0.42**	1.00													
0.77	**0.60**	**0.58**	1.00												
0.52	**0.35**	**0.57**	**0.61**	1.00											
0.71	**0.67**	**0.55**	**0.66**	**0.63**	1.00										
0.87	**0.73**	**0.77**	**0.85**	**0.76**	**0.86**	1.00									
0.19	**0.40**	**0.32**	**0.23**	0.13	*0.15*	**0.28**	1.00								
0.55	**0.46**	**0.46**	**0.51**	**0.53**	**0.47**	**0.60**	**0.29**	1.00							
0.59	**0.30**	**0.56**	**0.54**	**0.34**	**0.34**	**0.54**	**0.30**	**0.41**	1.00						
0.35	*0.19*	**0.38**	**0.25**	*0.21*	**0.43**	**0.36**	**0.23**	**0.27**	**0.34**	1.00					
0.41	**0.29**	**0.36**	**0.33**	**0.22**	**0.36**	**0.39**	**0.26**	**0.38**	**0.51**	**0.37**	1.00				
0.63	**0.46**	**0.64**	**0.56**	**0.45**	**0.51**	**0.64**	**0.52**	**0.67**	**0.76**	**0.59**	**0.76**	1.00			
0.86	**0.70**	**0.79**	**0.82**	**0.70**	**0.79**	**0.95**	**0.40**	**0.69**	**0.69**	**0.47**	**0.57**	**0.83**	1.00		
−0.75	**−0.56**	**−0.65**	**−0.69**	**−0.56**	**−0.69**	**−0.78**	**−0.30**	**−0.58**	**−0.56**	**−0.43**	**−0.41**	**−0.68**	**−0.83**	1.00	
−0.34	*−0.25*	**−0.44**	*−0.32*	**−0.39**	*−0.33*	**−0.42**	−0.06	**−0.60**	−0.23	*−0.29*	−0.16	**−0.47**	**−0.53**	**0.61**	1.00

coefficients[a]

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
120.															
120.	120.														
120.	120.	120.													
120.	120.	120.	120.												
120.	120.	120.	120.	120.											
120.	120.	120.	120.	120.	120.										
120.	120.	120.	120.	120.	120.	120.									
120.	120.	120.	120.	120.	120.	120.	120.								
120.	120.	120.	120.	120.	120.	120.	120.	120.							
120.	120.	120.	120.	120.	120.	120.	120.	120.	120.						
120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.					
120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.				
120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.			
120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	121.		
120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	120.	121.	121.	
45.	45.	45.	45.	45.	45.	45.	45.	45.	45.	45.	45.	45.	46.	46.	46.

Appendix J. Summary statistics and correlations of the court sample: summary statistics

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Variance-256	115.438	29.047	843.7	7.2618	16.
String-256	196.688	57.390	3293.6	14.3474	16.
Multiple	212.625	45.551	2074.9	11.3878	16.
Z score	35.313	3.092	9.6	0.7731	16.
Sex	1.063	0.250	0.1	0.0625	16.
Age	42.438	9.571	91.6	2.3926	16.
P	2.583	1.311	1.7	0.3786	12.
E	12.333	5.883	34.6	1.6982	12.
N	11.333	5.211	27.2	1.5042	12.
L	8.667	4.250	18.1	1.2268	12.
Variance-512	136.125	23.931	572.7	8.4609	8.
String-512	244.625	47.111	2219.4	16.6561	8.
Information	14.500	2.608	6.8	0.6519	16.
Comp	17.438	2.337	5.5	0.5843	16.

Appendix K. Summary statistics and correlations of the court sample: product-moment correlation

	1	2	3	4	5	6	7	8	9	10	11	12
1 Variance-256	1.00											
2 String-256	**–0.59**	1.00										
3 Multiple	–0.23	*0.48*	1.00									
4 Z score	*–0.48*	0.41	0.23	1.00								
5 Sex	–0.21	–0.01	–0.21	0.15	1.00							
6 Age	–0.26	*0.55*	0.29	0.10	0.32	1.00						
7 P	–0.09	–0.19	0.41	–0.15	–0.38	–0.19	1.00					
8 E	0.03	**–0.67**	–0.16	–0.13	0.41	–0.31	0.43	1.00				
9 N	**0.68**	–0.02	0.43	–0.19	–0.38	0.01	–0.04	–0.17	1.00			
10 L	–0.07	0.32	0.16	–0.05	–0.12	0.27	0.14	–0.06	0.11	1.00		
11 Variance-512	*0.78*	–0.20	0.45	–0.17	0.00	–0.03	–0.11	0.12	0.49	0.31	1.00	
12 String-512	–0.57	*0.76*	0.22	0.39	0.00	0.58	0.09	–0.44	–0.33	*0.68*	–0.17	1.00
13 Information	–0.27	*0.45*	*0.53*	*0.52*	–0.15	*0.51*	0.17	–0.13	0.43	0.10	0.23	0.14
14 Comp	–0.41	0.32	0.39	0.30	0.18	0.57	–0.14	–0.10	0.44	0.43	0.15	0.35
15 Arith	–0.30	0.34	0.23	–0.13	–0.37	*0.48*	0.43	–0.45	–0.10	0.19	–0.41	0.45
16 Simil	–0.13	*0.43*	*0.46*	0.42	–0.11	0.22	–0.11	–0.30	0.32	0.11	0.21	–0.02
17 Digit span	–0.36	*0.53*	0.19	*0.43*	–0.15	0.24	–0.08	–0.19	0.26	–0.40	–0.46	0.02
18 Vocabulary	–0.08	0.08	**0.59**	0.21	–0.31	–0.03	0.15	–0.27	*0.61*	–0.17	0.43	–0.34
19 Verb tot	–0.42	*0.54*	**0.59**	0.40	–0.18	*0.57*	0.04	–0.41	*0.50*	0.04	–0.02	0.19
20 Digit symb	–0.14	–0.24	–0.08	0.10	0.10	–0.21	0.49	*0.64*	–0.22	0.15	–0.13	–0.09
21 Pict comp	–0.35	0.42	**0.59**	0.39	–0.30	0.16	0.04	–0.49	0.32	–0.36	–0.41	–0.08
22 Blocks	**–0.71**	**0.81**	0.32	**0.59**	0.05	0.15	–0.13	–0.42	–0.33	0.10	–0.39	*0.72*
23 Pict arr	–0.25	0.29	*0.55*	0.04	–0.31	–0.06	*0.60*	0.11	0.00	–0.20	–0.17	–0.26
24 Obj assemb	*–0.57*	**0.72**	0.38	*0.44*	–0.13	0.04	0.27	–0.26	–0.17	0.18	–0.61	*0.68*
25 Perf tot	**–0.72**	**0.84**	**0.60**	*0.43*	–0.08	0.37	0.30	–0.27	–0.08	0.06	–0.53	0.53
26 IQ tot	**–0.66**	**0.80**	**0.67**	*0.48*	–0.14	*0.50*	0.19	–0.36	0.22	0.05	–0.32	0.39
27 Var-str 256	**0.80**	**–0.95**	*–0.44*	*–0.48*	–0.07	*–0.50*	0.12	*0.53*	0.25	–0.28	0.42	*–0.79*
28 Var-str 512	**0.81**	*–0.72*	0.01	–0.40	0.00	–0.50	–0.12	0.42	0.48	–0.44	0.56	**–0.91**

Appendix J. (continued)

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Arith	11.813	2.007	4.0	0.5018	16.
Simil	13.625	1.544	2.4	0.3860	16.
Digit span	12.813	3.125	9.8	0.7811	16.
Vocabulary	16.500	2.129	4.5	0.5323	16.
Verb tot	127.563	9.674	93.6	2.4186	16.
Digit symb	12.250	2.017	4.1	0.5041	16.
Pict comp	12.500	2.221	4.9	0.5553	16.
Blocks	12.438	3.119	9.7	0.7798	16.
Pict arr	10.563	3.032	9.2	0.7581	16.
Obj assemb	11.563	3.306	10.9	0.8265	16.
Perf tot	120.875	15.747	248.0	3.9369	16.
IQ tot	126.250	11.509	132.5	2.8774	16.
Var-str 256	–81.250	78.025	6087.9	19.5063	16.
Var-str 512	–108.500	56.277	3167.1	19.8971	8.

coefficients

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1.00															
0.64	1.00														
0.41	0.27	1.00													
0.56	0.27	0.34	1.00												
0.59	0.24	0.22	0.36	1.00											
0.53	*0.57*	0.15	0.34	0.13	1.00										
0.87	**0.78**	*0.56*	**0.60**	**0.62**	**0.62**	1.00									
–0.01	–0.29	–0.07	–0.03	–0.15	–0.39	–0.28	1.00								
0.41	0.35	0.23	*0.43*	*0.49*	**0.61**	**0.64**	–0.27	1.00							
0.31	0.15	0.04	0.26	*0.45*	0.11	0.33	–0.09	0.31	1.00						
0.23	0.06	0.29	0.26	*0.43*	0.37	0.35	–0.14	*0.50*	0.25	1.00					
0.17	–0.07	–0.04	0.16	*0.44*	–0.04	0.17	0.14	0.38	**0.86**	0.38	1.00				
0.48	0.26	0.29	0.33	**0.63**	0.16	*0.55*	0.07	**0.59**	**0.77**	*0.56*	**0.84**	1.00			
0.75	*0.55*	*0.46*	*0.52*	**0.72**	0.41	**0.85**	–0.09	**0.69**	**0.65**	*0.52*	**0.62**	**0.91**	1.00		
–0.43	–0.38	–0.36	–0.36	*–0.52*	–0.09	*–0.55*	0.12	*–0.44*	**–0.86**	–0.31	**–0.74**	**–0.88**	**–0.83**	1.00	
–0.02–	0.23	–0.56	0.10	–0.21	0.47	–0.17	0.02	–0.11	*–0.77*	0.15	**–0.82**	*–0.67*	–0.47	**0.84**	1.00

Appendix L. Summary statistics and correlations of the court sample: sample base of correlation

	1	2	3	4	5	6	7	8	9	10	11	12
7 P	12.	12.	12.	12.	12.	12.	12.					
8 E	12.	12.	12.	12.	12.	12.	12.	12.				
9 N	12.	12.	12.	12.	12.	12.	12.	12.	12.			
10 L	12.	12.	12.	12.	12.	12.	12.	12.	12.	12.		
11 Variance-512	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	
12 String-512	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.
13 Information	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
14 Comp	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
15 Arith	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
16 Simil	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
17 Digit span	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
18 Vocabulary	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
19 Verb tot	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
20 Digit symb	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
21 Pict comp	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
22 Blocks	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
23 Pict arr	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
24 Obj assemb	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
25 Perf tot	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
26 IQ tot	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
27 Var-str 256	16.	16.	16.	16.	16.	16.	12.	12.	12.	12.	8.	8.
28 Var-str 512	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.

coefficients[a]

13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
16.															
16.	16.														
16.	16.	16.													
16.	16.	16.	16.												
16.	16.	16.	16.	16.											
16.	16.	16.	16.	16.	16.										
16.	16.	16.	16.	16.	16.										
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
16.	16.	16.	16.	16.	16.	16.	16.								
8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.	8.

[a] Unless shown to the contrary, correlations are based on a sample size of 16.

Appendix M. Summary statistics and correlations for the mensa sample (Variable numbers are consistent with other subsamples)

1. Summary statistics

Variable	Mean	Standard Deviation	Variance	Standard Error	Number Cases
Variance-256	99.053	17.678	312.5	4.0555	19.
String-256	248.895	59.745	3569.4	13.7064	19.
Multiple	227.316	56.026	3138.9	12.8532	19.
Z score	33.105	3.999	16.0	0.9173	19.
Sex	1.632	0.496	0.2	0.1137	19.
Age	28.667	6.544	42.8	1.5424	18.
P	3.300	3.917	15.3	1.2387	10.
E	11.300	6.343	40.2	2.0058	10.
N	10.000	6.018	36.2	1.9032	10.
L	5.600	4.402	19.4	1.3920	10.
IQ tot	147.211	5.663	32.1	1.2991	19.
Var-str 256	–149.842	59.324	3519.4	13.6099	19.

2. Product-moment correlation coefficients

	1	2	3	4	5	6	7
1 Variance-256	1.00						
2 String-256	0.17	1.00					
3 Multiple	–0.27	0.17	1.00				
4 Z score	–0.21	0.27	–0.03	1.00			
5 Sex	0.15	0.33	–0.23	–0.04	1.00		
6 Age	*–0.52*	–0.35	–0.10	0.24	0.16	1.00	
7 P	*0.61*	0.25	0.04	0.38	*0.58*	–0.10	1.00
8 E	0.31	0.55	*0.57*	–0.12	–0.26	–0.34	0.07
9 N	*–0.57*	–0.40	0.45	0.23	–0.34	0.36	–0.09
10 L	–0.18	–0.05	–0.19	–0.17	–0.38	–0.00	**–0.76**
26 IQ tot	–0.09	–0.06	–0.12	0.32	0.23	0.15	0.44
27 Var-str 256	0.13	**–0.96**	–0.25	–0.33	–0.28	0.20	–0.05

3. Sample base of correlation coefficients (unless shown to the contrary (below), correlations are

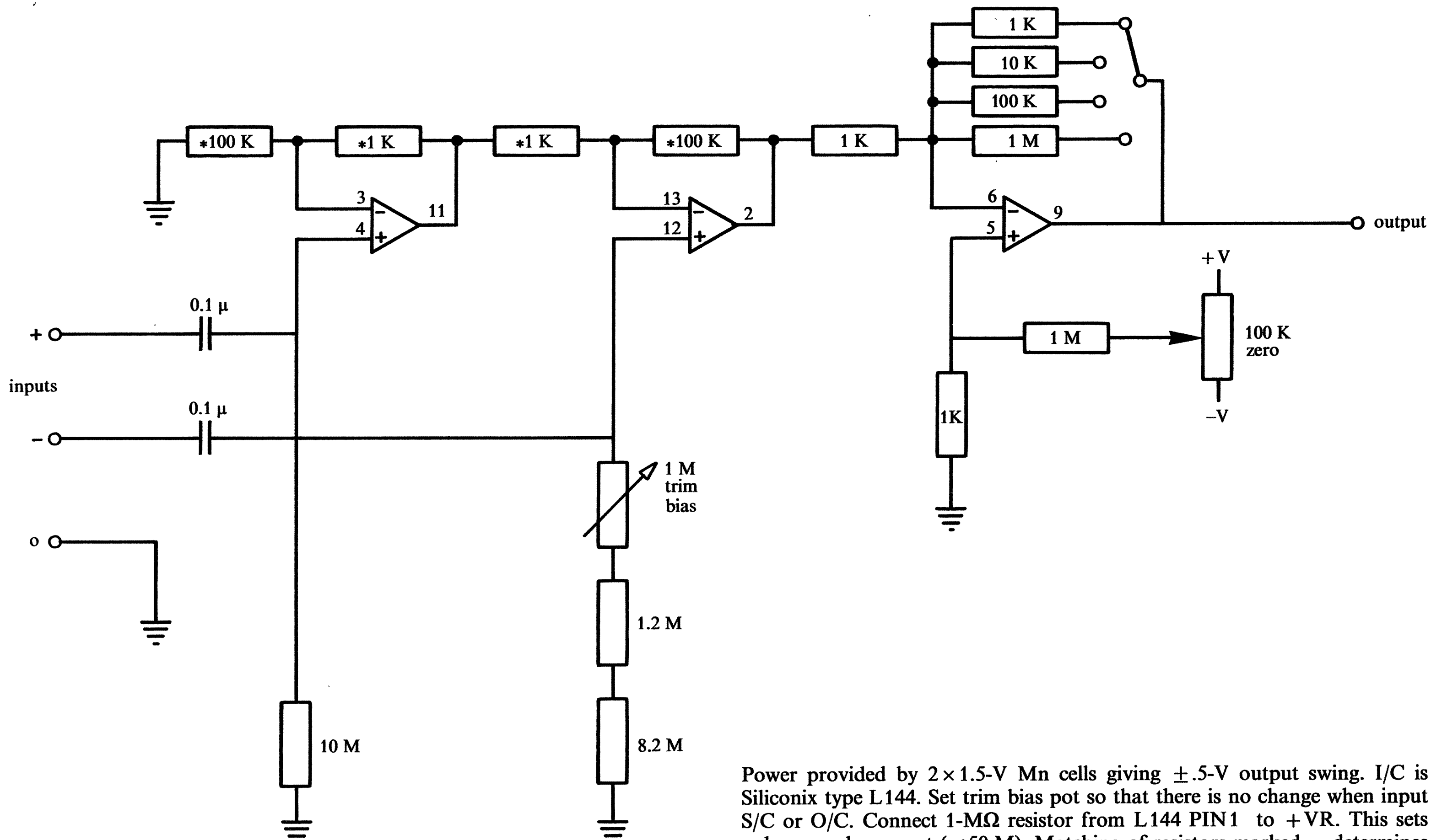
	1	2	3	4	5	6	7
6 Age	18.	18.	18.	18.	18.	18.	
7 P	10.	10.	10.	10.	10.	10.	10.
8 E	10.	10.	10.	10.	10.	10.	10.
9 N	10.	10.	10.	10.	10.	10.	10.
10 L	10.	10.	10.	10.	10.	10.	10.
26 IQ tot	19.	19.	19.	19.	19.	18.	10.
27 Var-str 256	19.	19.	19.	19.	19.	18.	10.

8	9	10	21	22	23	24	25	26	27
1.00									
0.36	1.00								
–0.08	–0.35	1.00							
–0.22	–0.11	–0.41	N/A	N/A	N/A	N/A	N/A	1.00	
–0.54	0.26	–0.01	N/A	N/A	N/A	N/A	N/A	0.03	1.00

based on a sample size of 19)

8	9	10	21	22	23	24	25	26	27
10.									
10.	10.								
10.	10.	10.							
10.	10.	10.	N/A	N/A	N/A	N/A	N/A	19.	
10.	10.	10.	N/A	N/A	N/A	N/A	N/A	19.	19.

Appendix N. EEG preamplifier (single integrated circuit; low battery consumption; ±1.5-V supply; P. Smith April 1977)



Power provided by 2 × 1.5-V Mn cells giving ±.5-V output swing. I/C is Siliconix type L144. Set trim bias pot so that there is no change when input S/C or O/C. Connect 1-MΩ resistor from L144 PIN1 to +VR. This sets v. low supply current (<50 M). Matching of resistors marked * determines CMRR. Take care to be accurate.

References

Berger H (1929) Über das Elektrenkephalogramm des Menschen. Arch Psychiatr Nervenkr 87:527–570

Blinkhorn SF, Hendrickson DE (1982) Averaged evoked responses and psychometric intelligence. Nature 295:596–597

Blomberg R, Cohen RS, Siekevitz P (1977) The structure of postsynaptic densities isolated from dog cerebral cortex. II. Characterisation and arrangement of some of the major proteins within the structure. J Cell Biol 74:204–225

Callaway E (1975) Brain electrical potentials and individual psychological differences. Grune & Stratton, New York

Cattell RB (1963) Theory of fluid and crystallized intelligence: a critical experiment. J Educ Psychol 54:1–22

Creutzfeldt OD, Rosina A, Ito M, Probst W (1969) Visual evoked response of single cells and of the EEG in primary visual area of the cat. J Neurophysiol 32:127–139

Ellingson RJ (1966) Relationship between EEG and test intelligence: a commentary. Psychol Bull 65:91–98

Ertl JP, Schafer EWP (1969) Brain response correlates of psychometric intelligence. Nature 223:421–422

Eysenck HJ, Eysenck SBG (1975) Manual of the Eysenck personality questionnaire. Hodder & Stoughton, London

Fox SS, O'Brien JH (1965) Duplication of evoked potential waveform by curve of probability of firing of a single cell. Science 147:838

Grab DJ, Carlin RK, Siekevitz PP (to be published) Ann NY Acad Sci

Grab DJ, Berzins K, Cohen RS, Siekevitz P (1979) J Biol Chem 254:8690–8696

Grundke-Iqbal I, Johnson AB, Wisniewski HM, Terry RD, Iqbal K (1979) Evidence that alzheimer neurofibrillary tangles originate from neurotubules. Lancet 17:578–580

Hendrickson AE (1972) An integrated molar/molecular model of the brain. Psychol Rep 30:343–368

Hendrickson DE (1972) An examination of individual differences in cortical evoked response, PhD thesis, University of London

Hendrickson DE, Hendrickson AE (1980) The biological basis of individual differences in intelligence. Pers Individ Diff 1:3–33

Hyams JS, Stebbings H (1979) Microtubule associated cytoplasmic transport. In: Roberts K, Hyams JS (eds) Microtubules. Academic Press, London, pp 487–530

Jarlfors V, Smith DS (1969) Association between synaptic vesicles and neurotubules. Nature 224:710–711

Jasper HH (1958) Report of the committee on methods of clinical examination in electroencephalography. EEG Clin Neurophysiol 10:370

Katz B (1966) Nerve, muscle, and synapse. McGraw-Hill, London

Kelly OT, Cotman CW (1977) Identification of glycoproteins and proteins at synapses in the central nervous system. J Biol Chem 252:786–793

Lin CT, Dedman JR, Brinkley BR, Means AR (1980) Localization of calmodulin in rat cerebellum by immunoelectron microscopy. J Cell Biol 85:473–480

Luduena RF (1979) Biochemistry of tubulin. In: Roberts K, Hyams JS (eds) Microtubules. Academic Press, London, pp 65–116

Matus A (1981) The postsynaptic density. TINS 4:51–53

Matus AI, Taff-Jones DH (1978) Morphology and molecular composition of isolated postsynaptic junctional structures. Proc R Soc Lond [Biol] 203:135–151

Matus AI, Walters BB (1975) Ultrastructure of the synaptic junctional lattice isolated from a mammalian brain. J Neurocytol 4:369–375

Matus AI, Walters BB, Mughal S (1975) Immunohistochemical demonstration of tubulin associated with microtubules and synaptic junctions in mammalian brain. J Neurocytol 4:733–744

Osborne RR (1970) Heritability estimates for the visual evoked response. Life Sci 9:481–490

Peters A, Palay SL, Webster HF (1970) The fine structure of the nervous system. The cells and their processes. Hoeber, New York

Saville P (1971) A British supplement to the manual of the Wechsler adult intelligence scale. Psychol Corp, New York

Scheele RB, Borisy GG (1979) In Vitro assembly of microtubules. In: Roberts K, Hyams JS (eds) Microtubules. Academic Press, London, pp 175–254

Shelanski ML (1973) Microtubules. In: Schneider DJ (ed) Proteins of the nervous system. Raven Press, New York, pp 227–242

Shelanski ML, Feit H, Berry RW, Daniels MP (1972) Some biochemical aspects of neurotubule and neurofilament proteins. In: Davison AN, Mandel P, Morgan IG (eds) Functional and structural proteins of the nervous system. Plenum Press, New York, pp 55–67

Shucard DW, Horn JL (1972) Evoked cortical potential and measurement of human abilities. J Comp Physiol Psychol 78:59–68

Taxi J (1967) Observations on the ultrastructure of the ganglionic neurons and synapses of the frog. Rana esculenta L. In: Hyden H (ed) The neuron. Elsevier, London, pp 221–252

Vaughan HG (1969) The relationship of brain activity to scalp recording of event-related potentials. In: Donchin E, Lindsley DB (eds) Av-

erage evoked potentials. NASA SP-191, Washington DC, p 45
Vogel W, Broverman DM (1964) Relationship between EEG and test intelligence: a critical review. Psychol Bull 62:132–144
Walters BB, Matus AI (1975) Tubulin in postsynaptic junctional lattice. Nature 257:496–498
Wechsler D (1955) Manual for the Wechsler adult intelligence scale. Psych Corp, New York
Westrum LE, Gray EG (1977) Microtubules associated with postsynaptic 'thickenings'. J Neurocytol 6:505–518
Wood JG, Wallace RW, Whitaker JN, Cheung WY (1980) Immunocytochemical localization of calmodulin and a heat-stable calmodulin-binding protein in basal ganglia of mouse brain. J Cell Biol 84:66–76

D Cognitive Principles and Intelligence

8 A Componential Interpretation of the General Factor in Human Intelligence

R.J. Sternberg and M.K. Gardner

Ever since Spearman's (1904, 1927) proposal of a general factor permeating all aspects of intelligent behavior, theorists of intelligence have busied themselves trying either to prove or disprove the existence in the mind of Spearman's "*g*". No doubt this popular pursuit will continue, if only because it provides a way of filling time for those who have had trouble finding other pursuits that strike their fancy.

We interpret the preponderance of the evidence as overwhelmingly supporting the existence of some kind of general factor in human intelligence. Indeed, we are unable to find any convincing evidence at all that militates against this view. We shall present here only a cursory examination of the main findings that lead us to accept the existence of a general factor, since careful and thorough reviews of the documentation exist elsewhere (e.g., Eysenck 1979, Humphreys 1979, McNemar 1964). For the most part, we shall assume that a general factor exists, and proceed to what we believe to be the interesting question facing contemporary theorists of intelligence: What is the nature of the general factor? In particular, we shall attempt to understand *g* in information-processing terms, applying a metatheoretical framework we refer to as a "componential" one in our attempt to isolate the information-processing origins of *g*. This framework has been used with at least some success in the analysis of a variety of different kinds of intelligent behavior (see Sternberg 1977b, 1978a, 1979, 1980c, 1980d, 1981a, 1981b). We certainly do not wish to claim that the componential framework is the only one in which general intelligence potentially can be understood: Any pie can be sliced in a number of ways, and the best we can hope for is that our way of slicing the pie yields pieces of a reasonable size and shape.

Our presentation is divided into five parts. First, we present a brief summary of some of the evidence that can be adduced in support of the existence of a general factor in human intelligence. Second, we present an overview of our beliefs regarding the nature of *g* as understood in componential terms. Third, we describe the research approach we use to tackle the problem of the nature of *g*, and state why we believe it is adequate to the problem, at least at one level of analysis. Fourth, we present evidence that supports our views regarding the nature of *g*. Fifth and finally, we summarize the main points of our argument.

Selected Evidence Supporting the Existence of General Intelligence

Various sorts of evidence have been adduced in support of the existence of general intelligence (Humphreys 1979). Perhaps the most persuasive evidence is everyday experience: Casual observation in everyday life suggests that some people are "generally" more intelligent than others. People's rank orderings of each other may differ according to how they define intelligence, but some rank ordering is usually possible. Moreover, when people are asked to characterize the behaviors that typify a "generally" intelligent person, they have no trouble in doing so, and there is a high degree of consistency both in the sorts of behaviors that are listed and in the perceived relationships among these behaviors, as ascertained by factor analysis (Sternberg et al. 1981). Very similar factor structures are obtained both for experts and laypersons: A generally intelligent person is conceived to be one who is particularly adept at the behaviors constituting problem solving, verbal facility, and common sense in interactions with the real world.

Historically, the evidence that has been offered most often in favor of the existence of general intelligence is the appearance of a general factor in unrotated factor solutions from factor analyses of tests of intelligence (e.g., Spearman 1927). Other factor-analytic techniques, such as second-order factoring of first-order factoring, can also yield a general factor. (See Jensen 1982, for a discussion of various factorial methods for eliciting a general factor.) In earlier research on the nature of mental abilities (e.g., Thurstone 1938), and in some contemporary research as well (e.g., Guilford 1967, Guilford and Hoepfner 1971), the general factor seems to disappear because of the way in which the factorial axes are rotated. For example, a general factor almost never appears when axes are rotated to Thurstonian "simple structure" (Thurstone 1947). But when correlated simple-structure factors are themselves factored, a general factor usually appears at the second order of analysis.

Many theorists of intelligence no longer view the debate over whether or not there is a general factor as still viable. Instead, they accept some kind of hierarchical structure of mental abilities whereby intelligence is viewed as comprising a general factor at the highest level, major group factors such as fluid and crystallized abilities (Cattell 1971, Horn 1968) or practical-mechanical and verbal-educational abilities (Vernon 1971) at the next level, minor group factors at a third level, and specific factors at a fourth level. What had seemed like conflicting views at one time, then, are now seen by these theorists, including ourselves, as basically compatible (Snow 1979, Sternberg 1980b, 1980d). Accepting this point of view, we can turn to the question of what kinds of entities generate individual differences in performance at the highest level of the hierarchy, that of general intelligence.

Were factor-analytic evidence the only kind that lent support to the existence of a general factor, one might write off the general factor as a method-specific peculiarity deriving somehow either from the mathematical mechanics of factor analysis or from the particular nature of individual-differences data. If one delves into the nature of variation across stimulus types rather than across subjects, however, a result parallel to the general factor emerges. A number of investigators, including ourselves, have used multiple regression techniques to isolate sources of stimulus variation in task performance. For example, we have attempted to predict response times to answer various kinds of analogies on the basis of manipulated sources of task difficulty in the solution of the analogies, e.g., the degree of relatedness between the first two terms, the degree of relatedness between the first and third terms, and so on (see Sternberg 1977a, b). A result that at first glance appears most peculiar has emerged from many of these task analyses (Egan 1976, Hunt et al. 1975, Jensen 1979, Keating and Bobbitt 1978, Mulholland et al. 1980, Sternberg 1977a, b): The regression intercept, or global "constant," often turns out to be as highly correlated or more highly correlated with scores from IQ tests than are the analyzed parameters representing separated sources of variance. Since the constant includes speed of response, e.g., button pressing, one could interpret such results trivially as indicating that motor speed is an essential ingredient of intelligence. A more plausible interpretation, and, as it will turn out, one more consistent with the bulk of the data, is that there are certain constancies in information-processing tasks that tend to be shared across wide variations in item types. We suggest that the search for the general component(s) and the search for the general factor are one and the same search – that whatever it is that leads to a unitary source of individual differences across subjects also leads to a unitary source of differences across stimulus types.

What is General Intelligence?

On the componential view, the basic construct underlying intelligent functioning is the information-processing component. A *component* is an elementary information process that operates upon internal representations of objects or symbols (Sternberg 1977b, see also Newell and Simon 1972). The component may translate a sensory input into a conceptual representation, transform one conceptual representation into another, or translate a conceptual representation into a motor output. What is considered elementary enough to be labeled a component depends upon the level of theorizing that is desired. Just as factors can be split into successively finer subfactors, so can components be split into successively finer subcomponents. Thus, no claim is made that any of the components referred to later are elementary at all levels of analysis. Rather, they are claimed to be elementary at a convenient level of analysis. The same caveat applies to the typology of components that will be proposed. Doubtless, other typologies could be proposed that would serve the present or other theoretical purposes as well or better. The particular typology proposed, however, has proved to be convenient in at least certain theoretical and experimental contexts. A number of theories have been proposed during the past decade that might be labeled, at least loosely, as componential (e.g., Butterfield and Belmont 1977, Campione and Brown 1979, Carroll 1976, 1980, Hunt 1978, Jensen 1979, Pellegrino and Glaser 1980, Snow 1979). The present theory, then, is just one of this general class of theories, although it is probably a bit more elaborated than at least some of the other theories.

Properties of Components

Each component has three important properties associated with it: *duration, difficulty* (i.e., probability of being executed erroneously), and *probability of execution.* Methods for estimating these properties of components are described in Sternberg (1978b) (see also Sternberg 1977b, 1980b, Sternberg and Rifkin 1979). It is dangerous to make inferences about one property of a component on the basis of information about another. We have found, for example, that the duration of a component is not necessarily correlated with its difficulty (Sternberg 1977a, b; 1980c).

Kinds of Components

Kinds of components can be classified in two different ways: by function and by level of generality.

Function. Components perform (at least) five kinds of functions. *Metacomponents* are higher-order control processes that are used for executive planning and decision-making in problem solving. *Performance components* are processes that are used in the execution of a problem-solving strategy. *Acquisition* (or storage) *components* are processes used in learning new information. *Retention* (or retrieval) *components* are processes used in retrieving previously stored knowledge. *Transfer components* are processes used in generalization, that is, in carrying over knowledge from one task or task context to another. Generally speaking, metacomponents act on other kinds of components (and on themselves), whereas performance, acquisition, retention, and transfer components act on information of various kinds.

Level of Generality. Components can be classified in terms of three levels of generality. *General components* are required for performance of all tasks within a given task universe. *Class components* are required for performance of a proper subset of tasks that includes at least two tasks within the task universe. *Specific components* are required for the performance of single tasks within the task universe. Tasks requiring intelligent performance differ in the numbers of components they require for completion and in the number of each kind of component they require.

Components and General Intelligence

To communicate early on the conclusion we will reach from an evaluation of the data we have collected, we assert here that individual differences in general intelligence can be attributed in part to individual differences in the effectiveness with which general components are performed. Since these components are common to all of the tasks in a given task universe, factor analyses will tend to lump these general sources of individual differences variance into a single general factor. As it happens, the metacomponents have a much higher proportion of general components among them than do any of the other kinds of components, presumably because the executive routines needed to plan, monitor, and possibly replan performance are highly overlapping across tasks of a widely differing nature. Thus, individual differences in metacomponential functioning are largely responsible for the persistent appearance of a general factor in mental-test data.

Metacomponents are probably not solely responsible for "*g*," however. Most behavior, and probably all of the behavior exhibited on intelligence tests, is learned. There may be certain acquisition components general across a wide variety of learning situations, which also enter into the general factor. Similarly, components of retention and transfer may also be common to large numbers of tasks. Finally, certain aspects of performance – such as encoding and response – are common to virtually all tasks, and they, too, may enter into the general factor. Therefore, although the metacomponents are primarily responsible for individual differences in general intelligence, they are almost certainly not solely responsible. Acquisition, transfer, retention, and performance components that are general across tasks also can be expected to contribute to individual differences in the general factor underlying intelligent performance.

In this second part of the chapter, we have given a very compact view of the nature of components and of how components enter into general intelligence. We proceed now to describe in some detail the methods of two as yet unpublished experiments addressed primarily to the question of what is general intelligence (Sternberg and Gardner 1980), and then describe more briefly other experiments upon which we shall draw that also address this question (Sternberg 1977a, Sternberg and Nigro 1980, Sternberg and Rifkin 1979).

Some Experimental Paradigms for Isolating the Information-Processing Origins of General Intelligence

We have conducted a number of experiments that have led us to the views described in the preceding part of the chapter. In terms of our present exposition, two particular experiments have been central to our conceptualizations, and several other experiments have been peripheral to these conceptualizations.

The "Central" Paradigm

The basic problem we confronted is that of isolating the information-processing origins of the general factor in human intelligence. Our basic strategy was to (a) select items that have been shown in the past to be excellent measures of *g*; (b) model response choices and response times in each of these items; (c) examine what emerged as common across the models and the tasks; and (d) propose an information-processing account of g on the basis of the observed communalities (Sternberg and Gardner 1980).

In most psychometric investigations of intelligence, the psychometric technique upon which the investigation has been based has been factor analysis. In such investigations, a representative sample of subjects from a population of interest would be given a range of tests sampling a wide variety of mental abilities, such as vocabulary, analogies, spatial visualization, classification, memory, and word fluency; then, an intercorrelation matrix would be computed between all possible pairs of these tests; next, the intercorrelation matrix would be factor analyzed to yield hypothesized latent sources of individual differences in the observable test scores; finally, interpretations would be assigned to these factors on the basis of the clusters of tests that showed high or low loadings on the various factors.

In our investigation of general intelligence, we also drew heavily upon a psychometric technique for analysis of the data. The technique we used was nonmetric multidimensional scaling rather than factor analysis, however (see Kruskal 1964a, b, Shepard 1962a, b, 1974). In our use of this technique, the goal was to discover the dimensions underlying a hypothetical semantic space comprising names of mammals, such as "lion," "tiger," "giraffe," "beaver," "donkey," and "rabbit." In a typical multidimensional-scaling study, subjects are asked to rate the similarity (or dissimilarity) between all possible pairs of terms to be scaled, which, in our case, were 30 mammal names. Next, a proximity matrix is formed comprising the mean rated similarity (or dissimilarity) of each term to every other term. It is usually assumed in advance that the matrix is reflexive (i.e., that the dissimilarity between a term and itself is zero), symmetrical (i.e., that the dissimilarity between one term and another is equal to the dissimilarity between the second term and the first), and that the triangle equality is satisfied (i.e., that if the distance between a first term and a second term is large, and the distance between that first term and a third term is large, then the distance between the second term and the third term is also large). Then, the multidimensional scaling algorithm is applied to the similarity or dissimilarity data, using only ordinal properties of the data, and yielding a psychological space comprising underlying dimensions of relationship among stimuli. Finally, the dimensions are interpreted on the basis of clusters of stimuli that have high or low loadings on each of the dimensions.

We were spared the need of actually doing the scaling ourselves by the fact that it had been done earlier on the set of mammal names by Henley (1969), who used a variety of different measures of relationship as input to the scaling algorithm and found striking consistencies in the outcome space without regard to the measure of relationship used. Henley found that the relations among mammal names could be captured very well by a three-dimensional spatial solution, with dimensions of size, ferocity, and humanness. For example, a gorilla would have a high loading on all three of these dimensions, whereas a beaver would have a low loading on all three. Henley used orthogonal dimensions in her solution, so that for the total set of mammal names, there was no correlation between loadings on pairs of dimensions.

We used the mammal names from the Henley (1969) scaling of proximity data to form 30 mammal-name analogies, series completions, and classifications. The analogies were taken from Rumelhart and Abrahamson's (1973) study of analogical reasoning with mammal names; the classifications and series completions were of our own construction. In Experiment 1, we administered each item untimed in four-choice, multiple-option format, with the subjects' task to rank-order each of the options in terms of

Table 1. Examples of problems used in mammal names reasoning experiments

Problem type	Experiment	
	Response choice	Response time
Analogy	TIGER:CHIMPANZEE::WOLF: (a. RACCOON, b. CAMEL, c. MONKEY, d. LEOPARD)	TIGER:CHIMPANZEE:: WOLF: (a. RACCOON, b. MONKEY)
Series completion	SQUIRREL:CHIPMUNK: (a. RACCOON, b. HORSE, c. DOG, d. CAMEL)	SQUIRREL:CHIPMUNK: (a. HORSE, b. CAMEL)
Classification	ZEBRA, GIRAFFE, GOAT, (a. DOG, b. COW, c. MOUSE, d. LEOPARD)	ZEBRA, GIRAFFE, GOAT, (a. MOUSE, b. LEOPARD)

its goodness of fit as a possible solution. In Experiment 2, we administered the same items, retaining just two of the four options; in this experiment, subjects were asked to select the better option as rapidly as they could. Examples of items are shown in Table 1. Subjects in the two experiments were 30 and 36 (different) college undergraduates respectively; obviously, our subject pool was not representative of the general population (in this or any of our experiments). Subjects received the three reasoning tasks in counterbalanced order, and then received a set of mental ability tests stressing reasoning abilities.

The "Peripheral" Paradigms

Sternberg (1977a) administered schematic-picture, verbal, and geometric analogies tachistoscopically to Stanford undergraduates. The first two kinds of analogies were presented in true-false format; the last kind was presented in forced-choice format. The analogies were standard in form (A:B::C:D, where D could be either a true or false completion or one of two answer options), and were easy enough to allow almost error-free performance in the subject population.

Sternberg and Nigro (1980) administered verbal analogies to 20 students in each of grades 3, 6, 9, and college. The college students were Yale undergraduates; the other students were public-school students from a middle-class suburb of New Haven. All subjects received the same 180 verbal analogies in which vocabulary level was restricted to grade 3 or below according to the Thorndike-Lorge norms. Analogies were presented in three formats differing in the numbers of terms in the analogy stem versus in the analogy options. Specifically, the number of terms in the analogy stem could be either three, two, or one. The remaining terms were options. Consider an example of each format: (a) Narrow: Wide:: Question: (trial) (statement) (answer) (task); (b) Win: Lose:: (dislike: hate) (ear: hear) (enjoy: like) (above: below); (c) Weak: (sick:: circle: shape) (strong:: poor: rich) (small:: garden: grow) (health:: solid: firm). Each option appeared on a separate line of print. Numbers of answer options varied from two to four. Items were presented tachistoscopically, and subjects were told to respond as quickly as possible.

Sternberg and Rifkin (1979) administered schematic-picture analogies to between 15 and 21 parochial-school children in each of grades 2, 4, and 6, and college-level adults at Yale. Analogies were presented in forced-choice format in 24 test booklets, each containing 16 analogies composed of binary attributes including height (tall, short), garment color (black, white), sex (male, female), and weight (fat, thin) (as in Sternberg 1977a). Items within each of the 24 booklets were homogeneous in terms of the number of attributes varied from the first term to the second, from the first term to the third, and between the two answer options. Since identities of actual values on attributes varied across analogies, however, no two analogies were identical. Each booklet was timed for 64 s. The main dependent variable, solution latency for items correctly answered, was computed by dividing 64 by the number of items correctly completed in a given booklet.

In an experiment on executive processing (see Sternberg 1981a) Salter and Sternberg administered to 20 Yale undergraduates verbal analogies that differed from standard analogies in that the positions of from one to three analogy terms could be occupied by multiple-choice options. The particular positions that were thus occupied differed from one item type to another. Either two or three alternative answer options were substituted for each missing analogy term (see also Lunzer 1965). An example of such a problem is Man: Skin::(dog, tree):(bark, cat). The correct answers are "tree" and "bark." The complete set of formats includes the following item types, where terms with the subscript i are missing ones with either two or three answer options substituted for the missing term: $A_i{:}B{::}C{:}D$; $A{:}B_i{::}C{:}D$; $A{:}B{::}C_i{:}D$; $A{:}B{::}C{:}D_i$; $A_i{:}B{::}C_i{:}D$; $A_i{:}B{::}C{:}D_i$; $A{:}B_i{::}C_i{:}D$; $A{:}B_i{::}C{:}D_i$; $A{:}B{::}C_i{:}D_i$; and $A{:}B_i{::}C_i{:}D_i$.

Item types from these peripheral paradigms, as well as those from the central paradigm, form the basis of the task analyses presented in the next part of the chapter.

Componential Investigations of General Intelligence

We have proposed a metatheoretical framework for theory construction in a recent chapter (Sternberg 1980d) that comprises a list of questions that a complete theory of intelligence ought at least to be able to address. We shall organize our discussion of our componential investigations of general intelligence around the questions proposed by this framework.

1. What Kind or Kinds of Problems Does the Theory Address?

Any attempt to provide an information-processing account of general intelligence (or any other kind of account) must start off with an appropriate set of tasks on the basis of which conclusions about general intelligence will be drawn. If the set of tasks is inappropriate, obviously, it does not matter much what kind of theorizing follows from it. In our approach, tasks are selected on the basis of four criteria originally proposed by Sternberg and Tulving (1977) in a different context and proposed in the present context by Sternberg (to be published b): quantifiability, reliability, construct validity, and empirical validity. The first criterion, quantifiability, assures the possibility of the "assignment of numerals to objects or events according to rules" (Stevens 1951, p. 1). The second criterion, reliability, measures true-score variation relative to total-score variation. In other words, it measures the extent to which a given set of data is systematic. The third criterion, construct validity, assures that the task has been chosen on the basis of some psychological theory. The theory thus dictates the choice of tasks, rather than the other way around. The fourth criterion, empirical validity, assures that the task serves the purpose in the theory that it is supposed to serve. Thus, whereas construct validity guarantees that the selection of a task is motivated by theory, empirical validity tests the extent to which the theory is empirically supportable.

Our choice of tasks in the investigation of general intelligence has included as its mainstays analogies, series completions, and classifications. The choice of these tasks was motivated largely by the criteria described above. First, performance on each of these tasks is readily quantifiable in terms of solution latency, error rate, response choice, and the like. Second, performance on these tasks has been reliably measured in countless tests of mental ability, as well as in a number of information-processing analyses of human intelligence. Third, the construct validity of these item types has been demonstrated in multiple ways. Factor analyses of intelligence-test batteries have shown these three kinds of items to be among those loading most highly on the general factor (see Cattell 1971, Guilford 1967, Guilford and Hoepfner 1971, Spearman 1927, Thurstone 1938). These tasks have played a central role in information-processing analyses of intelligence (see e.g., Evans 1968, Greeno 1978, Mulholland et al. 1980, Pellegrino and Glaser 1980, Simon 1976, Sternberg 1977b, 1979b, as well as in psychometric investigations; and they have even played an important role in Piagetian investigations (see e.g., Piaget 1972, Piaget et al. 1977). Indeed, the inclusion of these item types in so many theoretical investigations as well as practical measurements of intelligence strongly attests to their construct validity. Finally, the items have been shown in correlational analyses (usually presented in technical manuals for tests) to be highly correlated both with total scores on the test batteries in which they are contained and with external kinds of performance, such as school grades (see e.g., Cattell and Cattell 1963).

We make no claim that these are the only item types one might have chosen to study as an entrée to the general factor in intelligence, or even that they are the best item types to study. Another likely candidate, for example, is the matrix problem, which we interpret as consisting of multiple converging series completions presented in two dimensions (see e.g., Hunt 1974). We do believe, however, that our set of three tasks comprises an appropriate, although obviously incomplete, battery on the basis of which one may begin to analyze the general factor in human intelligence.

2. What Performance Components are Posited by the Theory?

A theory of general intelligence should state the performance components involved (either necessarily or optionally) in solution of the kinds of items dealt with by the theory. Investigators differ, of course, in where their ideas come from regarding the components used. They may do an implicit task analysis by going through a task themselves; they may use verbal reports supplied by subjects after testing; they may use think-aloud protocols supplied by subjects during test; or they may use their intuitions to expand or modify previous theories. Whatever their origin, the performance components should be specified and described.

The proposed theory posits use of up to seven performance components in the solution of analogies, series completions, and classification problems. The components are most easily explicated and their use in the task contexts shown by some examples of how they might be used in the solution of actual test problems as might be found on intelligence tests.

Consider as an example the analogy, Lawyer : Client : : Doctor : (a. medicine, b. patient). According to the theory, a subject *encodes* each term of the analogy, retrieving from semantic memory and placing in working memory attributes that are potentially relevant for analogy solution; next, the subject *infers* the relation between Lawyer and Client, recognizing, say, that a lawyer provides professional services to a client; then, the subject *maps* the higher-order relation between the first and second halves of the analogy, here recognizing that the first half of the analogy deals with the services of the legal profession and that the second half of the analogy deals with the services of the medical profession; next, the subject *applies* the relation inferred between the first two terms from the third analogy term, here, Doctor, to form an ideal point representing the ideal solution to the analogy; then, the subject *compares* answer options, seeking the ideal solution from among the answers presented; if none of the answer options corresponds to the ideal point, the subject must *justify* one of the answer options as preferable to the others, in that it is closest to the ideal point; in a rank-ordering task, multiple justifications may be needed as successive options are eliminated; finally, the subject *responds* with the chosen answer.

The same basic model can be extended to series completion problems. Consider, for example, the series completion, Truman:Eisenhower:(a. F. Roosevelt, b. Kennedy). The subject must *encode* each term of the series completion. Next, he or she *infers* the relation of succession between Truman and Eisenhower. Mapping is not necessary in this and other series problems, because there is no distinction between domain and range: All terms of the problem derive from a single, homogeneous domain, here, that of presidents of the United States. The subject must, however, *apply* the relation inferred between Truman and Eisenhower from Eisenhower to an ideal point, presumably, Kennedy. Next, the subject *compares* the answer options, seeking the one corresponding to the ideal point. If neither option (or in the case of more than two options, none of the options) corresponds to the ideal point, the subject *justifies* one option as closest to the ideal point. Suppose, for example, that option (b) was L. Johnson rather than Kennedy. This option would be preferable to F. Roosevelt, in that it names a successor to Eisenhower, but would be nonideal, in that it does not name an immediate successor. Finally, the subject *responds* with the chosen answer. As in the case of analogies, the rank-ordering task would require multiple justifications to determine which option is closest to the ideal point, of those options not yet ranked.

The model can also be extended to classification problems. Consider, for example, the problem, Nebraska, California, Vermont, (a. Texas, b. Reno). The subject must *encode* each term of the problem. Next, the subject must *infer* what is common to Nebraska, California, and Vermont, in essence seeking a prototype or centroid that abstracts what is common to the three terms; as was the case in the series completion problems, the subject need not map any higher-order relation, since all of the terms of the problem are from a single, homogeneous domain. In classification problems, application is also unnecessary, because the inferred centroid *is* the ideal point: The subject need not extrapolate in any way to seek some further ideal point. Next, the subject *compares* the answer options, seeking the ideal solution. If none is present, the subject *justifies* one option as closer to the ideal point than the other(s). Finally, the subject *responds*. As in the case of analogies and series completions, rank-ordering the options requires multiple executions of the *justification* component. Ranking in these problems and in the series completions proceeds according to a decision rule to be described.

The components of information processing in the three tasks are slightly different: The analogies task requires the full set of seven information-processing components; the series completion task requires a subset of six of the seven parameters in the analogies task; the classification task requires a subset of five of the six parameters in the series completion task. Thus, one would expect that for problems with terms of equal difficulty, analogies would be slightly more difficult than series completion problems, and series completion problems would be slightly more difficult than classification problems. In fact, mean latencies follow this predicted pattern.

The performance components described above are posited to be sufficient for describing the flow of information processing from the beginning to the end of task solution. Each contributes in some amount to the latency and difficulty of a given task item. In order to account for subjects' choices of response alternatives, it is necessary to supplement these components with a decision rule for option selection. The decision rule we use, following Rumelhart and

Abrahamson (1973), is Luce's (1959) choice axiom. We further propose, as did Rumelhart and Abrahamson, that relative rankings of answer options follow a negative exponential decay function, with the form of the decay function in part determined by the representation of information that is used. We shall describe our implementation of the rule further in the next section on representation.

3. Upon What Representation or Representations do These Components Act?

We doubt that there is any known test that is reasonably conclusive in distinguishing one form of representation from another. We therefore tend to assume our representations, and accept as indirect evidence supporting them the fits of process or response-choice models that are based upon these representations.

We believe that the form of representation a subject uses in solving a problem depends in part upon the content of the particular problem, and in part upon the subject's own preferences. In a standard item from an intelligence test, such as the analogy Washington:1::Lincoln:(a. 10, b. 5), for example, we believe subjects are likely to use an attribute-value representation. In such a representation, Washington might be encoded as [(president (first)), (portrait on currency (dollar)), (war hero (Revolutionary))], 1 might be encoded as [(counting number (one)), (ordinal position (first)), (amount (one unit))], Lincoln might be encoded as [(president (sixteenth)), (portrait on currency (five dollars)), (war hero (Civil))], and so on. The attribute-value representation can be extended to pictorial as well as verbal kinds of items. A black square inside a white circle, for example, might be represented as [((shape (square)), (position (surrounded)), ((color (black))), ((shape (circle)), (position (surrounding)), ((color (white)))].

In our joint research on mammal-name analogies, we have assumed the spatial representation of mammal names used by Henley (1969), Rips et al. (1973), and Rumelhart and Abrahamson (1973). The conceptual basis for the use of this representation in reasoning was first provided by these last investigators. Rumelhart and Abrahamson suggested that reasoning occurs when information retrieval depends upon the form of one or more relationships among words (or other units). Pursuing this definition of reasoning, these investigators claimed that probably the simplest possible reasoning task is the judgment of the similarity or dissimilarity between concepts. They assumed that the degree of similarity between concepts is not directly stored as such, but is instead derived from previously existing memory structures. Judged similarity between concepts is a simple function of the "psychological distance" between these concepts in the memory structure. The nature of this function and of the memory structure upon which it operates is clarified by their assumptions (after Henley 1969) that (a) the memory structure may be represented as a multidimensional Euclidean space and that (b) judged similarity is inversely related to distance in this space.

On this view, analogical reasoning (and, as we shall show, other forms of reasoning as well) may itself be considered to be a kind of similarity judgment, one in which not only the magnitude of the distance but also the direction is of importance. For example, we would ordinarily interpret the analogy problem, A:B::C:X_i, as stating that A is similar to B in exactly the same way that C is similar to X_i. According to the assumptions outlined above, we might reinterpret this analogy as saying that the directed or vector distance between A and B is exactly the same as the vector distance between C and X_i. The analogy is imprecise to the extent to which the two vector distances are not equal.

Rumelhart and Abrahamson formalized the assumptions of their model by stating that given an analogy problem of the form A:B::C:(X_1, X_2, ..., X_n), it is assumed that

A1. Corresponding to each element of the analogy problem there is a point in an m-dimensional space...

A2. For any analogy problem of the form A:B::C:?, there exists a concept I such that A:B::C:I and an ideal analogy point, denoted *I* such that *I* is located the same vector distance from *C* as *B* is from *A*. The coordinates of *I* are given by the ordered sequence $\{c_j+b_j-a_j\}, j=1$, m.

A3. The probability that any given alternative X_i is chosen as the best analogy solution from the set of alternatives $X_1, \ldots, X_n$ is a monotonic decreasing function of the absolute value of the distance between the point X_i and the point *I*, denoted $|X_i - \mathrm{I}|$. (p. 4)

The first assumption simply states that the concepts corresponding to the elements of the analogy exist and are locatable within the *m*-dimensional space representing the memory structure. The second assumption states that an ideal solution point also exists within the memory structure, and that this point also represents a concept; it is quite likely that the ideal point may not have a named mammal in the English (or any other) language. The third assumption states that the selection of a correct answer option is governed by the distance between the various answer options and the ideal point, such that less distant answer options are selected more often than are more distant answer options.

These assumptions permit ordinal predictions about the goodness of the various answer options, but do not permit quantitative predictions. In order to make quantitative predictions of response choices, Rumelhart and Abrahamson made assumption 3 more specific, and added two more assumptions:

3′. The probability that any given alternative X_i is chosen from the set of alternatives $X_1, \ldots, X_n$ is given by

$$\Pr(X_i | X_1, \ldots, X_n) = p_i = v(d_i)/[\sum_j^n v(d_j)],$$

where v( ) is a monotonically decreasing function of its argument.

4. $v(x) = \exp(-\alpha X)$, where X and α are positive numbers.

5. We assume that the subjects rank a set of alternatives by first choosing the rank 1 element according to 3′ and, then, of the remaining alternatives, deciding which is superior by application of 3′ to the remaining set and assigning that rank 2. This procedure is assumed to continue until all alternatives are ranked. (pp. 8–9)

The more specific version of assumption 3 (labeled 3′) is an adoption of Luce's (1959) choice rule to the choice situation in the analogy. Assumption 4 further specifies that the monotone decrease in the likelihood of choosing a particular answer option as best follows an exponential decay function with increasing distance from the ideal point. The model of response choice therefore requires a single parameter, α, representing the slope of the function. Rumelhart and Abrahamson actually had their subjects rank-order answer options. The investigators predicted the full set of rank orderings by assuming (in assumption 5) that once subjects had ranked one or more options, they would rank the remaining options in exactly the same way that they had ranked the previous options, except that they would ignore the previously ranked options in making their further rankings. Rumelhart and Abrahamson (1973) carried out three ingenious experiments that lent credence to their response-choice model of analogical reasoning.

We proposed a modest extension of the Rumelhart-Abrahamson model so that it could account for response choices in series completion and classification problems as well as in analogy problems. Figure 1 shows how the extended model accounts for response choices in each of the three types of problems.

Consider an analogy problem of the form, A:B::C:(D_1, D_2, D_3, D_4), where the subject's task is to rank-order the answer options in terms of how well their relation to C is parallel to that between B and A. In an analogy problem such as this one, the subject must find an ideal point, *I*, that is the same vector distance from *C* as *B* is from *A*. Having found this point, the

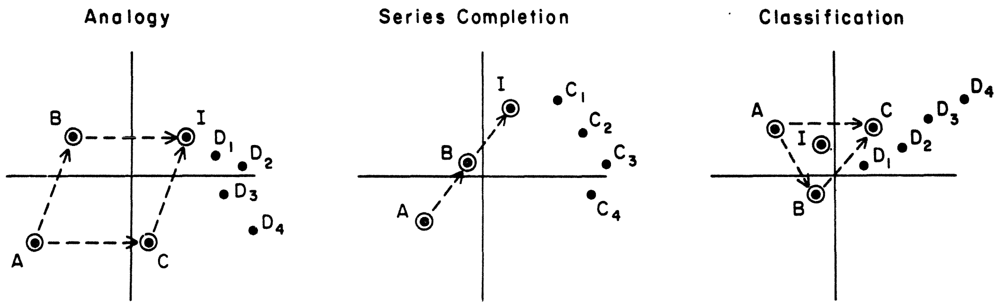


Fig. 1. Schematic diagrams showing rules for arriving at ideal point, *I*, in each of three induction tasks. In analogies, *I* is located as the fourth vertex in a parallelogram having *A*, *B*, and *C* as three given vertices. In series completions, *I* is located as the completion of a line segment that is at the same vector distance from *B* that *B* is from *A*. In classifications, *I* is the centroid of the triangle with *A*, *B*, and *C* as vertices. The rules can be extended to *n* dimensions by assuming *n*-dimensional analogues to the two-dimensional figures depicted. In each type of problem, options are presented at successively greater Euclidean distances from the ideal point

subject rank-orders answer options according to their overall Euclidean distance from the ideal point. The probability of selecting any one answer option as best is assumed to follow an exponential decay function, with probability decreasing as distance from the ideal point increases. The same selection rule is applied in rank-ordering successive options, with previously selected options removed from consideration.

Consider next a series completion problem of the form, A:B:(C_1, C_2, C_3, C_4), where the subject's task is to rank-order the answer options in terms of how well they complete the series carried from A to B. Here, the subject must find an ideal point, *I*, that is the same vector distance from *B* as *B* is from *A*. Note that the difference between a series completion problem and an analogy is that whereas the terms of an analogy form a parallelogram (or its *m*-dimensional analogue) in the multidimensional space, the terms of a series completion form a line segment (or its *m*-dimensional analogue) in the space. The same principle would apply, regardless of the number of terms in the item stem. Having found the ideal point, the subject rank-orders answer options with respect to the ideal point in just the same way that he or she would in an analogy problem.

Consider finally a classification problem of the form, A, B, C, (D_1, D_2, D_3, D_4), where the subject's task is to rank-order the answer options in terms of how well they fit with the three terms of the item stem. In this type of problem, the subject must find an ideal point, *I*, that represents the centroid in multidimensional space of *A*, *B*, and *C*. Having found this point, the subject rank-orders the answer options according to their overall Euclidean distance from the ideal point, in just the same way as he or she would for analogies or series completions. Again, the same basic principle applies without regard to the number of terms in the item stem. The centroid of the points is theorized always to serve as the ideal point.

Thus, we believe that the spatial representation can be used, at least in the context of terms falling into a semantic field, to represent information in a way that is suitable for the solution of three of the main types of problems used to measure general intelligence – analogies, series completions, and classifications.

4. By What Strategy or Strategies are the Components Combined?

Strategy refers to the order and mode in which components are executed. By "mode," we refer to whether the execution of a given set of components is serial or in parallel, exhaustive or self-terminating, and independent or nonindependent. In serial processing, components are executed sequentially; in parallel processing, they are executed simultaneously. In exhaustive processing, all possible executions of a given component or set of components are performed; in self-terminating processing, execution of components terminates before all possible executions have occurred. In independent processing, the execution of a given component has no effect upon whether any other component is executed; in dependent processing, execution of one component does affect whether one or more other components are executed.

In the Sternberg-Gardner experiments, we addressed the question of strategy only at a rather global level. The tests of the process model (in Experiment 2) were designed primarily to identify the components subjects actually used in solving the problems, rather than to identify how these components were combined. Our best evidence indicates that for the analogies, subjects would (a) encode the first term, (b) encode the second term, (c) infer the relation between the two terms, (c) encode the third term, (d) map the higher-order relation from the first half of the analogy to the second, (e) apply the previously inferred relation as mapped to the second half of the analogy to generate an ideal solution, (f) encode the two answer options, (g) compare the options, (h) justify one of the options as preferred, if nonideal, and (i) respond. For the series completions, we believe subjects would (a) encode the first term, (b) encode the second term, (c) infer the relation between the two terms, (d) apply the inferred relation to generate an ideal solution, (e) encode the two answer options, (f) compare the options, (g) justify one of the options as preferred, if nonideal, and (h) respond. For the classifications, subjects would (a) encode the first term, (b) encode the second term, (c) encode the third term, (d) infer the centroid, (e) compare the two answer options, (f) justify one of the options as preferred, if nonideal, and (g) respond.

More penetrating analyses of subjects' strategies were conducted in the analogical-reasoning experiments of Sternberg (1977a), Sternberg and Nigro (1980), and Sternberg and Rifkin (1979). These analyses enabled us to form detailed process models for the solution of each type of analogy. A flow chart representing the strategy most often used by adults for a wide variety of analogy types (schematic-picture, verbal, geometric) would show that subjects encode and infer as many attributes as they can find (exhaustive information processing), but map, apply, compare, and justify only a limited number of attributes (self-terminating processing). Subjects execute the self-terminating components in a self-terminating loop whereby they map, apply, and compare a single attribute at a time, seeking to disconfirm all but one answer option and then to justify one as acceptable; if the loop does not yield a satisfactory solution the first time around, it is iterated, this time with a second attribute. The process continues until it is possible to select one answer as the best of the given ones. Note that in this strategy, all components are assumed to be executed serially, and there is heavy process dependence in the sense that the outputs of earlier component executions are needed for later component executions. We have never actually compared serial versus parallel models of task performance, being convinced that the comparison is an extremely difficult one to carry out (see Pachella 1974).

5. What are the Durations, Difficulties, and Probabilities of Component Execution?

Table 2 shows parameter estimates for latencies of each component that was common to each of the three tasks studied in the Sternberg-Gardner experiments (except for inference, which was not statistically reliable in all three cases). If the three

Table 2. Parameter estimates for latency components

Parameter	Task	Parameter estimate[a]
Encoding	Analogies	1.22
	Series completions	1.00
	Classifications	.79
Comparison	Analogies	.13
	Series completions	.14
	Classifications	.14
Justification	Analogies	.36
	Series completions	.18
	Classifications	.24
Response +	Analogies	1.36
	Series completions	3.36
	Classifications	2.93

[a] Parameter estimates, expressed in seconds, are unstandardized linear regression coefficients. Comparison was estimated as a "time savings" for greater distance, but is expressed here in unsigned form. All coefficients are statistically significant at the 5% level or better.

tasks truly involve the same components of information processing, then the parameter estimates should be equal within a margin of error of estimation across tasks. A one-way analysis of variance was conducted across tasks upon each of the four parameter estimates of interest. Only the value of the justification parameter differed significantly across tasks (at the .001 level). Hence, the data are consistent with the notion that at least three of the components are common in kind across tasks, although obviously, further tests are needed. Justification could still be common across tasks but differentially difficult to execute, so that the existence of a significant difference does not totally refute the claim that the components are the same. Values of latency components differ, of course, with item content and format. We found, for example, that component latencies are generally lower for simple schematic-picture analogies than for simple verbal analogies, and lower for the verbal analogies than for geometric ones. What is of greatest interest is the relative amounts of time the various components consume. Encoding is always quite time-consuming, and the proportion of time it consumes is directly proportional to the complexity of the stimulus terms. The latency of response is about the same for different kinds of analogies, although the estimated parameter may differ as a function of other components that are sometimes confounded with response. (This confounding happens because response is estimated from the regression constant, which includes within it any source of latency that is common across all of the item types.) The amounts of time devoted to the other components vary greatly with analogy type, although it has been found that even small discrepancies between the ideal solution and the best of the given answer options can result in fairly substantial amounts of time spent in justifying this answer option as best, although nonideal (Sternberg 1977a, b).

Sternberg (1977a, b) predicted error rates as well as latencies for item solution. The finding of major interest was that self-terminating components were largely responsible for the errors that were made in item solution. In other words, the time saved by terminating information processing early is paid for in terms of the greater frequency of errors that are made due to what turns out to be premature termination of processing.

Sternberg and Gardner (1980) estimated α (the slope of the exponential function) as 2.52 for analogies, 2.56 for series completions, and 2.98 for classifications. Although these values did differ significantly from each other (due, obviously, to the higher value of α for the classification task), they are certainly in the same ballpark, and even the most extreme value corresponds roughly to that obtained by Rumelhart and Abrahamson for their analogies, 2.91.

The fits of the proposed theory to the various kinds of data were generally quite good in all of the experiments. In the Sternberg (1977a, b) experiments, values of R^2 between predicted and observed latencies were .92, .86, and .80 for schematic-picture, verbal, and geometric analogies, respectively. Values of R^2 were .85 and .89 respectively in the Sternberg-Nigro (verbal analogies) and Sternberg-Rifkin (schematic-picture analogies) experiments. And values of R^2 were .77, .67, and .61 for the analogies, series completions, and classifications in the Sternberg-Gardner experiment. For the model of response choice in this study, the values of R^2 were .94, .96, and .98 for analogies, series completions, and classifications, respectively.

6. What Metacomponents are Used in This Form of Information Processing?

We have proposed six metacomponents that we believe are critical in understanding intelligent information processing (Sternberg to be published a):

(1) Recognition of Just What the Problem is That Needs to be Solved. Anyone who has done research with young children knows that half the battle is getting the children to understand just what is being asked of them. Communication can also be a problem with adults, of course. Indeed, Resnick and Glaser (1976) have argued that intelligence is in large part the ability to learn in the absence of direct or complete instruction. Distractors on intelligence tests are frequently chosen so as to be the right answers to the wrong problems, so that they are chosen by those who do not recognize the problem that has been presented to them.

We found a rather striking example of the operation (or failure to operate) of this metacomponent in our developmental study with schematic-picture analogies (Sternberg and Rifkin 1979). In this experiment, certain second-graders consistently circled as correct one or the other of the first two analogy terms, rather than one or the other of the last two terms that constituted the answer options. We were puzzled by this systematic misunderstanding until we put together three facts – (a) that we were testing children in a Jewish parochial school, (b) that the children normally did their lessons in English in the morning and in Hebrew in the afternoon, and that (c) we happened to be doing our testing in the afternoon. Apparently, some of these young children perseverated in their normal afternoon right-to-left visual scanning, even in a task presented in English and where it was explicitly stated that the options were at the right. In the verbal analogies experiment of Sternberg and Nigro (1980), we also found a failure in the operation of this metacomponent: Some of the younger children (third and sixth graders) used association between words heavily in solving analogies, despite the fact that the task was presented as an analogical reasoning task.

(2) Selection of Lower-Order Components. An individual must select a set of lower-order (performance, acquisition, retention, or transfer) components to use in the solution of a given task. Selection of a nonoptimal set of components can result in incorrect or inefficient task performance. In some instances, choice of components will be partially attributable to differential availability or accessibility of various components. For example, young children may lack certain components that are necessary or desirable for the accomplishment of particular tasks, or may not yet execute these components in a way that is efficient enough to facilitate task solution. Two examples of changes in the selection of metacomponents with age come from our research on the development of analogical reasoning. First, we have found that young children (in Piagetian terms, those who are not yet formal-operational or even transitional into this period of development) do not map higher-order relations between the two halves of an analogy in their solution of analogy items. The mapping component is apparently either unavailable or inaccessible to such children (Sternberg and Rifkin 1979). Comparable results have been found by others as well (see e.g., Piaget et al. 1977). Second, we have found that whereas younger children

are quite prone to use an associative component in their solution of analogies, older children (those who are well into formal-operational thinking) do not (Sternberg and Nigro 1980). Again, these results are consistent with those of others (see e.g., Achenbach 1970, 1971).

(3) Selection of a Strategy for Combining Lower-Order Components. In itself, of course, a set of components is insufficient to perform a task: The components must be combined into a strategy. Strategy selection, like component selection, depends in part upon developmental level. In our developmental research on analogies, for example, we have found that children tend to modify their strategy for solving analogies as they grow older such that the strategy becomes increasingly more nearly exhaustive. The tendency to become more nearly exhaustive in information processing applies both within and between terms of analogies: Older children are more likely to encode as many attributes of each analogy term as they can and to infer as many relations between attributes of the first two analogy terms as they can than are younger children (Sternberg and Rifkin 1979); the older children are also more likely to search through all of the answer options in a given analogy, rather than choosing an answer as soon as they see an option that seems potentially appropriate (Sternberg and Nigro 1980).

(4) Selection of One or More Representations or Organizations for Information. A given component is often able to operate upon any one of a number of different possible representations or organizations for information. The choice of representation or organization can facilitate or impede the efficacy with which the component operates. In our research on the development of analogical reasoning, we have found evidence of changes in representation with age. Specifically, younger children are more likely to encode each of the attributes of a schematic-picture analogy separably, and then to make comparisons on each of the individual attributes; older children are more likely to integrate attributes and to treat the schematic pictures in a configural way (Sternberg and Rifkin 1979, Exp. 2). We have also found at least tentative evidence of individual differences in representations in adults. In our animal-name reasoning studies, we found that some individuals were more prone to use overlapping clusters of animal terms in addition to spatial dimensions than were others. For example, such a person might try to facilitate their analogy solution by realizing that animals like a tiger, lion, and panther are related in terms of dimensions such as size, ferocity, and humanness, but also in their all being jungle animals. Cats and dogs, on the other hand, are domesticated pets. But a household cat is related to the jungle animals by virtue of its being a feline animal, whereas a dog is not. The idea, then, is that animals are interrelated in a network of overlapping clusters that complements their dimensional attributes.

(5) Decision Regarding Allocation of Componential Resources. One of the barriers problem solvers encounter in solving problems is in the processing capacity they can bring to bear on a problem. Given that one's resources are limited, one must decide how many resources one can bring to bear on any given problem, given that there are usually competing demands for these resources. An example of differential resource allocation in action can be seen in our research on analogies with both children and adults. First, as children grow older, their latencies for analogy solution decrease. However, this composite latency can be decomposed into a series of component latencies that show that the global result is a gross oversimplification of what happens in analogy solution. It turns out that older subjects spend relatively *more* time than do the younger subjects in encoding the stimulus terms, but relatively *less* time in operating upon these encodings (Sternberg and Rifkin 1979, Exp. 1). Apparently, the older

children realize that obtaining a good fix on the nature of the stimulus later enables one to process that stimulus more efficiently, and thereby to save time, overall. Second, better adult reasoners solve analogy problems more quickly than do poorer adult reasoners. But this result, too, is an oversimplification. Complementary to the developmental finding is one that among adults, better reasoners tend to spend more time in encoding analogy terms than do poorer reasoners, but less time in operating upon these encodings (Sternberg 1977a, b). Thus, more sophisticated allocation of componential resources results in an overall improvement in performance.

(6) Solution Monitoring. As individuals proceed through a problem, they must keep track of what they have already done, what they are currently doing, and what they still need to do; the relative importances of these three items of information may differ across problems, but, nevertheless, all must be accomplished to some extent in every problem. That younger children are often less apt at solution monitoring than are older children is seen in the tendency of some of the second-graders in the pictorial analogies experiment to circle one of the two analogy terms at the left rather than the right of the problem (Sternberg and Rifkin 1979). Almost all of the second-graders were able to solve most analogies successfully, given that they understood what to do. The insensitivity of these subjects to the fact that right-to-left solution almost never yielded a suitable solution, much less, a suitable analogy, can be viewed as a failure of these subjects to monitor their solution processes adequately.

Even very young children do monitor their solutions to some extent, however. The use of solution monitoring in even the reasoning of very young children can be seen in the metacomponential decision of children of as young as the third-grade level to use a justification component in the solution of verbal analogies. The component continues to be used throughout the age span to adulthood. This performance component is elicited upon the recognition by a subject that none of the presented answer options in a multiple-choice analogy provides an ideal completion for the given problem. In such an event, the subject may have to justify one of the presented options as nonideal, but superior to the alternative options. The justification component is something of a "catchall," in that it includes in its latency any reexecution of previously executed performance components that may be attempted in an effort to see whether a mistaken intermediate result has been responsible for the subject's failure to find an optimal solution. The decision to use this component reflects an awareness on the part of the subject that things are not going quite right: The path to solution has reached a dead end, and some route must be found that will yield an ideal answer, or else an answer must be selected that is acceptable, if nonideal.

7. What are the Effects of (a) Problem Format, (b) Problem Content, and (c) Practice upon Intellectual Performance?

All of these variables have effects upon intellectual performance, at least in reasoning by analogy. Consider, for example, the effect of true-false versus multiple-choice format. In true-false analogies, solution can be quite simple if analogies are essentially digital in character, by which we mean that an answer is clearly either right or wrong. In schematic-picture analogies, for example, specific attributes such as height, clothing color, sex, and weight of pictures of people might be manipulated: The correct answer would be one that had the appropriate values on each of these four attributes. Suppose that one is asked instead to solve verbal analogies, however. It is actually quite rare that any given fourth term will be precisely correct; indeed, it is not even clear what "precisely correct" means for verbal analogies. For example, is Happy:Sad ::Tall:Small a true analogy or a false one? Usually, Short rather than Small is contrasted antonymously to Tall. Whereas Small does not seem quite right, it does not quite seem wrong either. Or consider the analogy, Car:Gas::Person:Food. Obviously, the two lower-order relationships (between Car and Gas and between Person and Food) that comprise this analogy are parallel in some ways, but not in others. On what basis could one say whether the analogy is "true" or "false," however? In multiple-choice analogies, the situation is different. On the one hand, one's task is complicated by the fact that it is now necessary to eliminate several incorrect options, some of which may be quite close to the best answer, rather than merely to indicate whether a given answer is correct or not; on the other hand, one's task is to choose the *best* answer, not the *right* answer. One can select an option knowing full well that it is not right or ideal in any meaningful sense of these terms, but that it is the best of the options that have been presented. The sources of difficulty are thus changed considerably when one moves from true-false to forced-choice analogical reasoning (Sternberg 1977b).

The effects of problem content are sometimes hard to predict in advance. Sternberg (1977a, b), for example, found that people handle verbal and geometric analogies in surprisingly similar ways. Sternberg and Rifkin (1979), however, found that two kinds of schematic-picture analogies that on their face look quite similar are processed in quite different ways. Analogies with clearly separable attributes are processed with maximum self-termination by adult subjects; analogies with attributes that are integral are processed with a combination of self-terminating and exhaustive information-processing components.

Consider finally the effects of practice upon analogy solution. Sternberg (1977b) compared performance during a first session of schematic-picture analogy solution to performance during a fourth (and final) session. As would be expected, latencies and error rates decreased from the first session to the fourth. All components showed shorter latencies during the fourth session than during the first except for inference. There was no evidence of strategy change across sessions: Fits of the various models and variants of models were almost identical in the two different sessions. The most interesting difference showed up during external validation of scores: In the first session, no correlations of latencies for the analogy items with scores on reasoning tests were significant; in the fourth session, more than half of the correlations were significant, and many of them were of high magnitude, reaching into the .60s and .70s. Results such as these led Glaser (1967) to conclude that psychometric test scores are more highly correlated with performance after asymptote is reached than with performance during initial trials of practice.

8. What are the Salient Sources of Individual Differences in Intellectual Performance at a Given Age Level?

The major loci of individual differences in intellectual performance in the componential approach to intelligence reside in the various kinds of components of human intelligence. Each component of each kind potentially can generate individual differences in performance. Sternberg (1977b) found substantial individual differences in the speeds at which the various performance components of analogical reasoning are executed, and in the degree to which subjects used any systematic strategy at all. No substantial individual differences were found in components or forms of representation used (although Sternberg & Gardner did find evidence of such representational differences). In terms of strategy differences, the main source of variation was that some adults seemed to be self-terminating in their inference process, although most were apparently exhaustive in this process.

9. What are the Salient Sources of Individual Differences in Intellectual Performance Across Age Levels (i.e., in Intellectual Development)?

We believe that the most important sources of developmental differences are metacomponential ones. Indeed, the section on metacomponents (No. 6) showed developmental trends in all of the metacomponents considered. On this view, the major source of development is in executive planning and decision making in problem solving. We have also found developmental differences in rates and accuracies of component execution (e.g., Sternberg 1979a, 1980a, Sternberg and Nigro 1980, Sternberg and Rifkin 1979). But the significance of these changes for development seems much smaller than the significance of the metacomponential changes, and indeed, we believe that these differences are attributable in large part to metacomponential changes. More efficacious planning and decision-making enable problem solvers to become more rapid and accurate in their problem solving. Consider, for example, the large decrease in error rates that has been observed in our developmental studies of analogical reasoning. Earlier analyses (Sternberg 1977b) had shown that errors in analogy solution were due almost entirely to premature self-termination of information processing. This finding, coupled with the finding that children become more nearly exhaustive in their information processing with increasing age, suggests that the tendency to become more nearly exhaustive may account at least in part for the developmental decrease in error rates that is observed.

10. Relationships Between Components of Various Intellectual Tasks

Individual parameter estimates were not reliable in the Sternberg-Gardner study, so it was not feasible to intercorrelate them. Intercorrelations were computed, however, between mean response latencies for subjects for each pair of data sets: The correlations were .85 between analogies and series completions, .86 between analogies and classifications, and .88 between series completions and classifications. A principal-components factor analysis of the three sets of latencies revealed a strong general factor in the individual-differences data, with the first, unrotated principal component accounting for 91% of the variance in the data. Had the tests shown no overlap in individual-differences variation (zero intercorrelations), this factor would have accounted for only 33% of the variation. The data are thus consistent with the notion that a single real-time information-processing model might apply across tasks.

A comparable set of analyses was performed on the ability-test scores: Here, the correlations were .72 between analogies and series completions, .45 between analogies and classifications, and .65 between series completions and classifications. A principal-components factor analysis of the three sets of test scores (numbers correct) revealed an unrotated, general first factor accounting for 74% of the variance in the individual-differences data. Again, such a factor would have accounted for only 33% of the variation had the tasks been unrelated. These results, too, therefore, are consistent with the notion of common processes across tasks.

Finally, intercorrelations were computed between task scores across the two forms of task presentation (tachistoscopic, leading to response latencies, and pencil-and-paper, leading to numbers correct). Correlations across task format were lower than those within format, as would be expected if there were at least some medium-specific variance that were not shared across task formats. Such medium-specific variance might result from differences across task formats in speed-accuracy trade-offs, in attentional allocations for items presented singly (as in a tachistoscopic task) and for items presented as a group (as in a pencil-and-paper task), in kinds of strategy or other planning required, or in what is measured by latency and accuracy scores. The correlations ranged from −.21 to −.41, with a median for the nine intertask correlations of −.35 ($P<.05$). Correlations of tasks with their analogues across formats (e.g., tachistoscopic analogies with pencil-and-paper analogies were only trivially higher than correlations of nonanalogous tasks across formats (e.g., tachistoscopic analogies with pencil-and-paper series completions): The median correlation for analogous tasks was −.35 ($P<.05$), whereas the median correlation for nonanalogous tasks was −.30 ($P<.05$). A factor analysis of the six tasks (three tachistoscopic and three pencil-and-paper) yielded a first, unrotated principal component accounting for 57% of the variance in the data. If tests were unrelated, a value of 17% would have been expected. As expected, the second unrotated principal component, accounting for 26% of the variance in the data, was a bipolar factor distinguishing pencil-and-paper tasks from response-latency ones. The general factor unifying the various kinds of tasks was thus about twice as strong as the medium-specific factor differentiating the two task formats. Subsequent factors were of little interest.

11. What are the Relationships Between the Components of the Set of Intellectual Tasks of Interest and General Intelligence?

Sternberg (1977b) found that each of the major components in analogical reasoning – inference, mapping, application, justification – can correlate with performance on tests of general intelligence when the attributes of the analogies being solved are nonobvious. As would be expected, faster latencies were associated with higher test performance. The latency of the response component was also very highly correlated with IQ test scores, although this finding was given a metacomponential interpretation: Metacomponents constant across the item types were at least partly responsible for the high correlation between the regression constant and the test scores (see Sternberg 1979b). Finally, encoding was also correlated with test scores, but in the opposite direction (as mentioned earlier): Slower encoding was associated with higher reasoning abilities. This finding, too, was interpreted metacomponentially as indicating a strategy whereby slower encoding was associated with faster operations upon the better encodings that resulted, so that overall performance was facilitated. Many of these findings have since been replicated (e.g., Mulholland et al. 1980).

12. What are the Practical Implications of What We Know About the Forms of Intellectual Behavior Covered by the Given Theory?

We have devised a training program for the metacomponents and performance components described earlier that we hope to implement in the near future (see Sternberg to be published b). To date, we have done research only on training the performance components of analogical reasoning (see Sternberg et al. 1982). We have found that it is possible to train people to use various different strategies for solving analogies, and that strategy training can greatly reduce correlations between component latencies and measured intelligence.

Sternberg (1977b) has argued that inductive reasoning such as that measured by series completions, classifications, and especially analogies is pervasive in everyday experience. "We reason analogically whenever we make a decision about something new in our experience by drawing a parallel to something old in our experience. When we buy a new pet hamster because we liked our old one or when we listen to a friend's advice because it was correct once before, we are reasoning analogically" (p. 99).

Oppenheimer (1956) has pointed out the signal importance of analogy in scientific reasoning of the kind done by scientists and even nonscientists on an everyday basis:

> Whether or not we talk of discovery or of invention, analogy is inevitable in human thought, because we come to new things in science with what equipment we have, which is how we have learned to think, and above all how we have learned to think about the relatedness of things. We cannot, coming into something new, deal with it except on the basis of the familiar and old-fashioned. The conservatism of scientific enquiry is not an arbitrary thing; it is the freight with which we operate; it is the only equipment we have. (pp. 129–130)

Analogical reasoning also plays an important role in legal thinking, where it may be called "reasoning by example" (Levi 1949):

> The basic pattern of legal reasoning is reasoning by example. It is reasoning from case to case. It is a three-step process described by the doctrine of precedent in which a proposition descriptive of the first case is made into a rule of law and then applied to a next similar situation. The steps are these: similarity is seen between cases; next the rule of law inherent in the first case is announced; then the rule of law is made applicable to the second case. This is a method of reasoning necessary for the law, but it has charcteristics which under other circumstances might be considered imperfections. (pp. 1–2)

Consider, in general, how the metatheoretical framework described in this chapter might be applied to diagnostic and prescriptive problems in educational and everyday theory and practice.

Suppose we know that a certain child is a poor reasoner. We might know this because of the child's low scores on psychometric tests of reasoning ability or because the child performs poorly in school on problems requiring various kinds of reasoning. The kinds of analyses suggested here yield a number of indices for each child (or adult) that can help localize the source of difficulty. These sources correspond to the basic sources of individual differences described earlier. One can discover whether certain components needed to solve one or more kinds of intellectual problems are unavailable, or available but not accessed when needed; whether the child is using a suboptimal strategy, that is, one that is time-consuming, inaccurate, or unable to yield any solution at all; whether the child finds execution of certain components especially difficult or time-consuming; whether the child is inconsistent in his or her use of strategy; or whether the child fails in metacomponential decision-making about problem solution. This information can then be used to prescribe the kind of remediation needed by the child.

Summary

To summarize, general intelligence can be understood componentially as deriving in part from the execution of general components in information processing behavior. Most general components are metacomponents, although performance, acquisition, retention, and transfer components also can be general in nature. Metacomponents dominate the information-processing system because they are the source of all direct activation of other kinds of components and because only they receive direct feedback from other kinds of components, as well as among themselves.

Componential metatheory requires a theory of general intelligence to deal with twelve questions about the nature of intelligence and its interaction with the real world. These questions were posed, and answers were given based on the research we and our colleagues have done using various componential techniques. The proposed theory was able at least to provide tentative answers to all of these questions.

We wish to emphasize in closing that we know, as should others, that our account of general intelligence is limited to one level of analysis, and is incomplete in many respects. We believe, for example, that the functioning of general intelligence in the real world cannot be understood completely without reference to the motivational variables that drive intellectual functioning, and hence that any account of general intelligence that is wholely cognitive (as is ours) cannot account for all of the behavioral patterns that we can reasonably label as "generally intelligent" (see also Zigler 1971). Hence, we present our account as one step toward a more all-encompassing theory that will view intelligence in all of its multifarious aspects.

Acknowledgment. Preparation of this chapter was supported by Contract N0001478C0025 from the Office of Naval Research to Robert J. Sternberg.

References

Achenbach TM (1970) The children's associative responding test: a possible alternative to group IQ tests. J Educ Psychol 61:340–348

Achenbach TM (1971) The children's associative responding test: a two-year followup. Dev Psychol 5:477–483

Butterfield EC, Belmont JM (1977) Assessing and improving the executive cognitive functions of mentally retarded people. In: Bialer I, Sternlicht M (eds) Psychological issues in mental retardation. Psychological Dimensions, New York, pp 277–318

Campione JC, Brown AL (1979) Toward a theory of intelligence: contributions from research with retarded children. In: Sternberg RJ, Detterman DK (eds) Human intelligence: perspectives on its theory and measurement. Ablex, Norwood NJ, pp 139–164

Carroll JB (1976) Psychometric tests as cognitive tasks: a new "structure of intellect." In: Resnick LB (ed) The nature of intelligence. Erlbaum, Hillsdale NJ, pp 27–56

Carroll JB (1980) Individual difference relations in psychometric and experimental cognitive tasks. NR 150–406 ONR technical report: final report. Thurstone Psychometric Laboratory University of North Carolina, Chapel Hill NC

Cattell RB (1971) Abilities: their structure, growth, and action. Houghton Mifflin, Boston

Cattell RB, Cattell AKS (1963) Test of g: culture fair, scale 3. Institute for Personality and Ability Testing, Champaign IL

Egan DE (1976) Accuracy and latency scores as measures of spatial information processing. Research report 1224. Naval Aerospace Medical Research Laboratories, Pensacola

Evans TG (1968) A program for the solution of geometric-analogy intelligence test questions. In: Minsky M (ed) Semantic information processing. MIT Press, Cambridge MA. pp 271–353

Eysenck HJ (1979) The structure and measurement of intelligence. Springer, Berlin Heidelberg New York

Glaser R (1967) Some implications of previous work on learning and individual differences. In: Gagné RM (ed) Learning and individual differences. Merrill, Columbus

Greeno JB (1978) Natures of problem-solving abilities. In: Estes WK (ed) Handbook of learning and cognitive processes, vol 5. Erlbaum, Hillsdale NJ, pp 239–270

Guilford JP (1971) The nature of human intelligence. McGraw-Hill, New York

Guildord JP, Hoepfner R (1971) The analysis of intelligence. McGraw-Hill, New York

Henley NM (1969) A psychological study of the semantics of animal terms. J Verb Learn Verb Behav 8:176–184

Horn JL (1968) Organization of abilities and the development of intelligence. Psychol Rev 75:242–259

Humphreys LB (1979) The construct of general intelligence. Intelligence 3:105–120

Hunt EB (1974) Quote the raven? Nevermore! In: Gregg LW (ed) Knowledge and cognition. Erlbaum, Hillsdale NJ, pp 129–158

Hunt EB (1978) Mechanics of verbal ability. Psychol Rev 85:109–130

Hunt EB, Lunneborg C, Lewis J (1975) What does it mean to be high verbal? Cognit Psychol 7:194–227

Jensen AR (1979) *g*: Outmoded theory or unconquered frontier? Creat Sci Technol 2:16–29

Jensen AR (1982) The chronometry of intelligence. In: Sternberg RJ (ed) Advances in the psychology of human intelligence, vol 1. Erlbaum, Hillsdale NJ

Keating DP, Bobbitt BL (1978) Individual and developmental differences in cognitive-processing components of mental ability. Child Dev 49:155–167

Kruskal JB (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27

Kruskal JB (1964b) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:28–42

Levi EH (1974) An introduction to legal reasoning. University of Chicago Press, Chicago

Luce RD (1959) Individual choice behavior. Wiley, New York

Lunzer EA (1965) Problems of formal reasoning in test situations. In: Mussen PH (ed) European research in cognitive development. Monogr Soc Res Child Dev 30:19–46

McNemar Q (1964) Lost: our intelligence? Why? Am Psychol 19:871–882

Mulholland TM, Pellegrino JW, Glaser R (1980) Components of geometric analogy solution. Cognit Psychol 12:252–284

Newell A, Simon HA (1972) Human problem solving. Prentice-Hall, Englewood Cliffs NJ

Oppenheimer JR (1956) Analogy in science. Am Psychol 11:127–135

Pachella RG (1974) The interpretation of reaction time in information processing research. In: Kantowitz B (ed) Human information processing: tutorials in performance and cognition. Erlbaum, Hillsdale NJ, pp 41–82

Pellegrino JW, Glaser R (1980) Components of inductive reasoning. In: Snow RE, Federico P-A, Montague W (eds) Aptitude, learning, and instruction: cognitive process analyses of aptitude (vol 1) Erlbaum, Hillsdale NJ, pp 177–218

Piaget J (1972) The psychology of intelligence. Littlefield Adams, Totowa NJ

Piaget J, Montangero J, Billeter J (1977) Les correlats. L'abstraction réfléchissante. Presses Universitaires de France, Paris

Resnick LB, Glaser R (1976) Problem solving and intelligence. In: Resnick LB (ed) The nature of intelligence. Erlbaum, Hillsdale NJ, pp 205–230

Rips L, Shoben E, Smith E (1973) Semantic distance and the verification of semantic relations. J Verb Learn Verb Behav 12:1–20

Rumelhart DE, Abrahamson AA (1973) A model for analogical reasoning. Cognit Psychol 5:1–28

Shepard RN (1962a) The analysis of proximities: multidimensional scaling with an unknown distance function. I. Psychometrika 27:125–140

Shepard RN (1962b) The analysis of proximities: multidimensional scaling with an unknown distance function. II. Psychometrika 27:219–246

Shepard RN (1974) Representation of structure in similarity data: problems and prospects. Psychometrika 39:373–421

Simon HA (1976) Identifying basic abilities underlying intelligent performance of complex tasks. In: Resnick L (ed) The nature of intelligence. Erlbaum, Hillsdale NJ, pp 65–98

Snow RE (1979) Theory and method for research on aptitude processes. In: Sternberg RJ, Detterman DK (eds) Human intelligence: perspectives on its theory and measurement. Ablex, Norwood NJ, pp 105–138

Spearman C (1904) General intelligence, objectively determined and measured. Am J Psychol 15:201–293

Spearman C (1927) The abilities of man. Macmillan, London

Sternberg RJ (1977a) Component processes in analogical reasoning. Psychol Rev 84:353–378

Sternberg RJ (1977b) Intelligence, information processing, and analogical reasoning: the componential analysis of human abilities. Erlbaum, Hillsdale NJ

Sternberg RJ (1978a) Componential investigations of human intelligence. In: Lesgold A, Pellegrino J, Fokkema S, Glaser R (eds) Cognitive psychology and instruction. Plenum, New York, pp 277–298

Sternberg RJ (1978b) Isolating the components of intelligence. Intelligence 2:117–128

Sternberg RJ (1979a) Developmental patterns in the encoding and combinaton of logical connectives. J Exp Child Psychol 28:469–498

Sternberg RJ (1979b) The nature of mental abilities. Am Psychol 34:214–230

Sternberg RJ (1980a) The development of linear syllogistic reasoning. J Exp Child Psychol 29:340–356

Sternberg RJ (1980b) Factor theories of intelligence are all right almost. Educ Res 9:6–13, 18

Sternberg RJ (1980c) Representation and process in linear syllogistic reasoning. J Exp Psychol [Gen] 109:119–159

Sternberg RJ (1980d) Sketch of a componential subtheory of human intelligence. Behav Brain Sci 3:573–584

Sternberg RJ (1981a) Intelligence and nonentrenchment. J Educ Psychol 73:1–16

Sternberg RJ (1981b) Toward a unified componential theory of human intelligence: I. Fluid abilities. In: Friedman M, Das J, O'Connor N (eds) Intelligence and learning. Plenum, New York, pp 327–344

Sternberg RJ (to be published a) Instrumental and componential approaches to the training of intelligence. In: Chipman S, Segal J, Glaser R (eds) Thinking and learning skills: Current research and open questions (vol 1). Erlbaum, Hillsdale New York

Sternberg RJ (to be published b) Reasoning, problem solving, and intelligence. In: Sternberg RJ (ed) Handbook of human intelligence. Cambridge University Press, New York

Sternberg RJ, Gardner MK (1980) Unities in inductive reasoning. NR150–412 ONR technol report 18. Department of Psychology Yale University, New Haven

Sternberg RJ, Nigro G (1980) Developmental patterns in the solution of verbal analogies. Child Dev 51:27–38

Sternberg RJ, Rifkin B (1979) The development of analogical reasoning processes. J Exp Child Psychol 27:195–232

Sternberg RJ, Tulving E (1977) The measurement of subjective organization in free recall. Psychol Bull 84:539–556

Sternberg RJ, Conway BE, Ketron JL, Bernstein M (1981) People's conceptions of intelligence. J Personality and Soc Psychol: Attitudes and Social Cognition 41:37–55

Sternberg RJ, Ketron JL, Powell JS (1982) Componential approaches to the training of intelligent performance. In: Detterman DK, Sternberg RJ (eds) How and how much can intelligence be increased? Ablex Norwood, New Jersey

Stevens SS (1951) Mathematics, measurement and psychophysics. In: Stevens SS (ed) Handbook of experimental psychology. Wiley, New York, pp 1–49

Thurstone LL (1938) Primary mental abilities. University of Chicago Press, Chicago

Thurstone LL (1947) Multiple factor analysis. University of Chicago Press, Chicago

Vërnon PE (1971) The structure of human abilities. Methuen, London

Zigler E (1971) The retarded child as a whole person. In: Adams HE, Boardman III WK (eds) Advances in experimental clinical psychology, vol 1. Pergamon Press, New York, pp 47–121

9 Epilogue: Is Intelligence?

H.J. Eysenck

We have now arrived at the end of our survey of novel developments in the theory and measurement of intelligence.

It will be clear to the reader that the discovery of a physiological measure and basis for IQ differences has many important consequences, not only for theory, but also for experimental design and investigation. It should now be possible to investigate a number of problems, which have hitherto been very resistant to scientific solution, by means of the R measure. A few examples must suffice to indicate some of the avenues newly opened up. Among the most important of these is the temporal course of the development of intelligence, from babyhood through childhood and adolescence to adulthood and senility. IQ tests cannot properly be applied until the age of 5 or 6, thus missing what are often assumed to be the most formative years. EEG measures of the evoked potential should be feasible at a much earlier age, making it possible to trace the development of intelligence from year 1 right through to death. Furthermore, the possibility exists that we may be dealing with a measure that has a true zero point, and has genuinely additive features unlike the IQ, which is meaningful only as a deviation from population standards, not in absolute terms. These possibilities of course will require much detailed study, but the transformation of a mentalistic to a physicalistic type of test holds out exciting possibilities along these lines.

In a similar fashion, we should be able to trace the decline of intelligence with age in a much more rigorous fashion than has been possible hitherto. As is well known, the multiplicity of IQ tests show very varying rates of decline with age, depending on such factors as time allowed, g_f vs g_c content, motivation, etc. (Eysenck 1979). By substituting a simple physiological measure of R, we could tap the biological factors underlying IQ performance, unadulterated by cultural, educational, and socioeconomic factors. Such studies should throw much-needed light on this problem, which is getting more important every day as the mean age of the population is increasing due to improved medical care, better hygiene, and advances in the nutritional sciences.

A third possibility, already mentioned, is the investigation of hemispheric differences in evoked potential measures related to verbal and non-verbal abilities. If it is true that verbal abilities are more closely related to the left hemisphere, non-verbal ones to the right, then individuals showing marked discrepancies on the two Wechsler scales say, should show corresponding differences in their R scores derived from the left and the right hemispheres, respectively. Unfortunately the long-term V-P discrepancy seems to be so unstable that the hope of finding a biological basis for it may be remote (Yule et al. to be published), but cognitive tasks more directly bearing on the topic may be available which bring out the discrepancy more strongly and more lastingly.

Yet another area of interest is the question of sex differences. There would seem considerable agreement that male-female differences in IQ are minimal, although on some first-order factors females would seem to do better (e.g. verbal), while on others males do better (e.g. visuospatial). However, the existence of such differences in first-order factors implies that IQ differences depend crucially on the combination of subtests selected for the purpose, and while the possibility exists that selection of subtests has not been independent of the desire to eliminate sex differences in the total test score, it will be desirable to use a biological measure such as the EEG R score, in order to eliminate any possibility of bias. The same consideration applies to the possibility that females show lower variances on IQ tests than do males, a difference attributed by Lehrke (1978) to sex linkage. The evidence is complex, and relies much on intrafamilial correlations for IQ. Thus sex linkage would lead one to expect that correlations of test scores for mother-daughter, father-daughter, and mother-son would be somewhat similar, parent and child in each case having one X-chromosome in common. The correlations between fathers and sons should be lower since they have no X-chromosomes in common, and the brother-sister correlations should be intermediate since they have an X-chromosome in common half the time. Results quoted by Lehrke support this hypothesis, but clearly it would be most desirable to replicate the studies quoted, using the evoked potential measure instead of IQ tests.

DE Hendrickson in her chapter has already discussed some of the findings regarding the close agreement between male and female EEG intelligence scores, and has drawn attention to the fact that as in the case of Wechsler IQ (and other traditional IQ measures) the variance of females is smaller than that of males on the AEP. We may use the variance score (which she regards as the most satisfactory of all those investigated) to look at an argument which might with advantage be extended to other samples, larger in size than that used in her investigation. As Eysenck (1979) has argued, something like 80% of the variance on the traditional IQ measures is accounted for by genetic causes, 20% by environmental causes. If we now look at differences between boys and girls in terms of variance, we would consider that the environmental contribution would lessen the disparity between the sexes, in view of the similar and possibly identical type of early education which they are subjected to. Thus we would expect that the differences in variance would be less on the WISC than on the AEP variance measure, and indeed this is so – in the Hendrickson data, the disparity is 6.4% for the WISC, and 16.7% for the AEP variance measure. It should, however, be noted that for the string measure the discrepancy, although in the same direction, is very much smaller; it is not known why there should be such a difference in the results of the two measures which are highly correlated, and are presumed to measure much the same underlying factors. Further work along these lines is clearly indicated.

It may be helpful to state the simple problem which most research in this general field has encountered, and the reason why most of the resulting difficulties do not arise in connection with the R measures. We are for the most part concerned with differences (old versus young, male versus female, psychotic versus normal, high verbal versus high performance score, etc) and their causation. The causes may be genetic or environmental, and the causal analysis is bedevilled by the fact that IQ measures combine both these causal factors in the proportion of roughly 80%–20% of total variance accounted for (Eysenck 1979). Some workers in the field put the proportions somewhat differently, but for the purpose of this discussion the actual proportions in question are immaterial. Now in most cases the observed differences are too small to allow an unequivocal judgement concerning the causes involved; hence the endless disputations and discussions between environmentalists and hereditarians. Ingenious experimental paradigms have been suggested and used (as pointed out in connection with the Lehrke study), but the resulting designs are often difficult to carry out, and it would clearly be preferable to have a relatively pure measure of genotypic intelligence, not influenced to any appreciable degree by education, cultural, or socioeconomic factors. It is our belief that R is such a measure, and hence can be used to improve our understanding of the complex type of problem here under discussion.

Such a belief of course requires substantiation. The high correlation with Wechsler and Matrices IQ indicates that the measures are concerned essentially with the same underlying (latent) trait, namely intelligence, and the obvious lack of cultural and educational features in the determination of R suggests strongly that R deviates from IQ in the direction of lesser dependence on such environmental factors. One way of supporting this argument would be by genetic studies of R, using twins, or any of the other methods of biometrical genetical analysis (Eysenck 1979). Another method is perhaps more interesting, and has the added advantage of simultaneously dealing with a substantive problem, namely the causes of the frequently observed differences in IQ between the children of high SES parents and low SES parents. These differences have been attributed to genetic causes, cultural-educational causes, and combinations of the two. It seems that the use of R can decisively improve experimental designs for the study of this problem.

Let us consider two groups of children, taken from the Hendricksons' study reported in an earlier chapter, on the basis of their high or low socioeconomic status, respectively. Both groups were given the WISC, as a measure of traditional IQ, and the evoked potential was ascertained, giving the two measures which are taken together to form the R score, i.e. the 'string' or complexity of the trace, and the variance obtained over 90 testings. The former correlates positively with IQ, the latter negatively. Now on the WISC the two groups had total IQ scores of 120.40 and 97.12 respectively, giving a difference of 23.28 points. The total group of children had a S.D. of 13.91; thus, dividing the difference into the S.D. gives us, in standard terms, a value of 1.67 – the difference between the two groups is 1.67 times the value of the S.D. If we wished to test the hypothesis that the string and the variance measure on the evoked potential were relatively 'pure' measures of genetic intelligence, what would be our prediction concerning the differentiation of the two groups of children on these measures?

It is known that the total variance of a test like the WISC is made up of roughly 80% genetic and 20% environmental factors; consequently of the observed difference of 1.67 S.D., 20% would be due to environmental factors, and only 80% to genetic factors. But if the evoked potential measures could be regarded as almost entirely genetic in origin, as far as individual differences are concerned, then the observed differences should be 20% less than for the WISC, because of the absence of the environmental factors, i.e. in terms of the S.D. of the means involved, the difference should be 1.34 S.D.s instead of 1.67 S.D.s. The results are shown in Table 1; it will be seen that the string measure is almost exactly right (1.33), and the variance measure is in excess of the expected value, being 1.18. These two values are not significantly different from each other, or from the predicted value, suggesting that the results support the hypothesis. The fact that they exceed the target, even if not significantly, suggests that perhaps the value of 80% is a trifle too high, or that the reliabilities of the tests are not identical. Figure 1 illustrates the results.

Table 1. Means and standard deviations on three tests of high SES and low SES children

Test		High SES (N=25)	Low SES (N=25)	Diff.	S.D.	Diff. in standard terms
WISC IQ		120.40	97.12	23.28	13.91	1.67
String (Complexity)	EEG	173.44	101.92	71.52	53.83	1.33
Variance	EEG	133.60	198.80	65.20	55.00	1.18

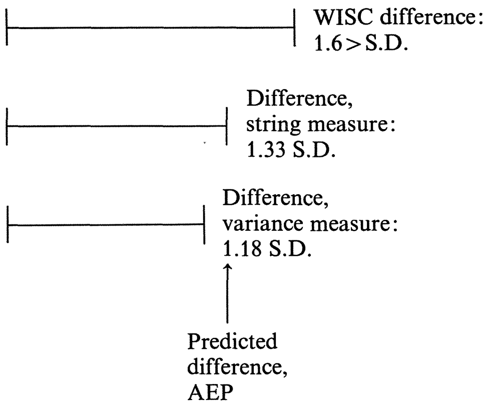


Fig. 1. Differences between high and low SES groups of children on WISC IQ, and on two EEG measures of evoked potential

These calculations are not presented as proof that the EEG measures are in fact pure measures of 'intelligence A' (genotypic intelligence); they are presented as suggestions of the kind of experiment that now becomes possible in order to check deductions from hypotheses such as these. Obviously the numbers involved are not large enough to make the results conclusive; far larger samples will be required in order to arrive at more persuasive results. Altogether, replication of the results reported in this book will be needed on a large scale to bolster up the suggestive investigations reported here. But already there is some evidence from other laboratories (e.g. Salzburg), or from the reanalysis of previously published work, like that of Ertl, to suggest that the data here reported are replicable, provided only that proper safeguards are taken to make the measurements conform properly as far as details of experimental design are concerned – the results are very much influenced by even quite small departures from optimal design, and in our own work we too have had failures due to neglect of certain experimental details which only later experience showed up as crucial (e.g. Rust, 1975)[1]. The whole history of the physiological measurement of intelligence is dotted with failure to replicate; it is only now that we are beginning to know just what are the crucial variables requiring control. Even such variables as the alloy used for the electrodes, or the precise definition of the stimulus intensity and make-up, can be crucial; replications must be exact in order to deserve the name! (see chapter by E. Hendrickson).

The claims made for this book are not that it records final achievements, but rather that it opens new doors, and suggests novel ways of attacking old problems. If it achieves that aim, we shall be well satisfied. The air was getting pretty stuffy as far as IQ measurement was concerned, and the arguments too passionate to find solutions along the old ways. It is our hope that the new directions indicated for future research may serve to let in some fresh air, and enable greater agreement to be reached on doubtful issues. Only the future will tell whether this hope is justified.

1 It may be useful to spell out some of the details which differentiated the Rust experiment from the Hendrickson experiment, and account for the failure of the former, and the success of the latter. Rust (1975) used stimuli of 95 dB, as contrasted with Hendrickson's 80 dB; he used 20 stimulus presentations as compared to her 90; and he used regular intervals of 33 s as compared to the irregular intervals used by her. Averages of 20 presentations give very different waveforms from those obtained by averaging 90 presentations, but more importantly a 33-s interval with 95-dB stimulations produces what amounts to a near startle response with all of its associated artefacts. The mean latencies and amplitudes obtained by Rust are so different from those obtained by Hendrickson that these stimulus differences must be considered very significant indeed. To say this is not to imply any criticism of the work of Rust; prior to trying out different parameter values it would not have been possible to say which procedure was in fact optimal for obtaining the best results. The point is an important one as Kamin (1981) quotes Rust's paper as evidence that evoked potentials do not correlate with intelligence. Both Rust and Hendrickson, in fact, worked in my Department in part to resolve this question of optimal parameter values, and the negative results of the one should not be used to throw doubts on the positive results achieved by the other.

Do the results reported in this volume enable us to give any sort of answer to the question which forms the title of this epilogue: Is intelligence? It would be difficult to deny that the results are not compatible with a model such as Guilford's which eliminates the general factor of intelligence completely. As far as they go, they suggest strongly the existence of a fundamental biological property of the CNS, underlying success on orthodox IQ tests as well as speed on RT measures, success on Inspection Time experiments, and of course errorless information processing through the CNS, with the latter presumably being the most fundamental biological correlate (and probably cause) of intelligence B (intelligent behaviour in ordinary life situations). This interpretation may not be correct, of course, but no alternative suggests itself at the moment, and as far as it goes this hypothesis does seem to account reasonably well for the observed phenomena.

We thus find it difficult to reject the view that intelligence is and remains a useful concept scientifically: a concept, moreover, which is firmly tied to physiological measurement and laboratory experimentation. This conclusion agrees well with the results of most psychometric investigations, which time and again come up with powerful evidence for a general factor which accounts for far more of the variance than all other factors combined. The possibility exists of course that both parts of this argument could be mistaken, but their strong agreement on all essentials does suggest that it would be premature to throw overboard the notion of intelligence as a fundamental concept in the analysis of problem solving behaviour. Our results of course also suggest that the psychometric IQ is less monolithic than it is usually assumed to be, and that much further experimental analysis is needed to disclose its dimensionality, and the relation of these dimensions to the biological aspects we have been discussing. Fortunately all these tasks are ideally suited to the processes of Kuhn's ordinary science, i.e. detailed working out of a paradigm which is agreed to be of value. It is our hope that such working out will extend the model here presented, and will establish it even more firmly as a fundamental contribution to the analysis of cognitive behaviour.

References

Eysenck HJ (1979) The structure and measurement of intelligence. Springer, Berlin Heidelberg New York

Kamin L (1981) In: Eysenck HJ, Kamin L (eds) The intelligence controversy. Wiley, New York

Lehrke RG (1978) Sex linkage: a biological basis for greater male variability in intelligence. In: Osborne, RT, Noble CE, Weyl N (eds) Human variation. Academic Press, New York

Rust J (1975) Cortical evoked potential, personality and intelligence. J Comp Physiol Psychol 89:1220–1226

Yule W, Gold R, Busch C (to be published) Long-term predictive validity of the WPPSI: an eleven-year follow-up study

Author Index

Subject Index

A Model for Personality

Editor: H. J. Eysenck
With contributions by numerous experts

1981. 75 figures. XII, 287 pages
ISBN 3-540-10318-X

Contents: Introduction. – General Features of the Model. – The Psychophysiology of Extraversion and Neuroticism. – A Survey of the Effects of Brain Lesions upon Personality. – The Genetic and Environmental Architecture of Psychoticism, Extraversion and Neuroticism. – Personality and Conditioning. – Learning, Memory and Personality. – Personality and Social Behaviour. – A Critique of Eysenck's Theory of Personality. – Epiloque. – Subject Index.

It is suggested in this book, that the time has come for the very electric and almost random type of research in personality to be supplanted by a paradigm which would permit normal scientific methods to be applied in this field. Eysenck and his co-authors suggest that such a paradigm exists, and each chapter reviews a different aspect of it – the extraversion/introversion dimension, genetic factors in personality, the relationship between personality and conditioning, memory, social behaviour, the effects of brain lesions and drugs on personality – with a special chapter devoted to alternative approaches.
It is concluded that there is a considerable amount of experimental evidence which converges on a model for personality fulfilling the scientific demands for a proper paradigm.

Springer-Verlag
Berlin
Heidelberg
New York