J. Social Biol. Struct. 1980 3, 103–122

Chronometric analysis of intelligence

Arthur R. Jensen
Institute of Human Learning,
University of California, Berkely, USA

Measurements of various parameters derived from different reaction time (RT) paradigms are found to be correlated with psychometric measurements of general mental ability. Such RT-derived measurements, when combined in a multiple regression equation, predict some 50 per cent or more of the variance in IQ or *g*. This relationship of IQ or *g* to RT parameters indicates that our standard IQ tests tap fundamental processes involved in individual differences in intellectual ability and not merely differences in specific knowledge, acquired skills, or cultural background.

Introduction

This article reviews the main currents in research on the relationship of reaction time (RT) to general intelligence and other psychometric metal abilities. In preparation I have read everything I could find in the literature directly relating RT to mental ability, from Galton to the present. (The first bonafide study of RT and intelligence that is reported completely enough to be technically evaluated is a study done at Yale University by J. A. Gilbert in 1894.) The literature on RT and intelligence is not massive; my stack of reprints – virtually the entire literature – is barely four inches high. This is only a small fraction of the research on strictly experimental, parametric studies of RT, which treats both inter- and intra-individual variability as mere nuisance variables. Trying to draw generalizations from a literature that is so spread over time, with so little uniformity of measurement techniques or methods of analysis, and that is seldom guided by any systematic theoretical conceptions is a bit like looking at a Rorschach inkblot. I can only hope that my perception of the most salient features of the available evidence is not too greatly at variance with the conclusions of other reviewers. I claim no more than to state what I have come to believe at present from my reading of this literature, including my own studies of RT in relation to psychometric abilities. I am not bothered, at this stage, that one can find exceptions to almost every generalization one could make. This should be expected where sampling error is considerable, where experimental effects and their correlations with other variables are generally small, where relatively small sample sizes are the rule because RT studies usually involve lengthy individual testing in the laboratory, where samples are usually much more homogeneous in ability than the general population,

A glossary of technical terms is provided in an appendix.

0140-1750/80/020103+20 $02.00/0

and where the temporal stability of certain RT measurements is much lower than is typically found for standard psychometric tests. To garner any generalizations that are of theoretical interest and are worthy of further investigation, one has no choice but to trust one's own scientific judgment to detect the key signals through the noise of procedural variations and sampling errors.

Correlation between RT and psychometric ability

The first and most general conclusion we can draw with a great deal of confidence is that measurements of RT parameters in a variety of paradigms are indeed significantly related to scores on standard tests of intelligence and other psychometric abilities. As I have noted elsewhere (Jensen, 1979), the study of RT as a measure of mental ability got off to a bad start in the early history of psychology, for a number of reasons, largely due to psychometric naiveté and inadequate statistical methods. Modern investigators have been more successful in finding substantial and replicable relationships between RT and IQ. I say 'relationships' at this point, rather than *correlations*, because not all relationships are expressed as correlation coefficients.

Correlation coefficients between RT and IQ are not as impressive or as consistent as are mean differences in RT between different criterion groups selected on the basis of IQ or other psychometric indices of ability. Correlations between RT and IQ can be generally characterized as fairly low. But in the entire literature on RT and IQ there are virtually no correlations on the 'wrong' side of zero. Most *r*s fall in the range from 0 to $-0{\cdot}50$, with a mode in the $-0{\cdot}30$'s. A correlation of $-0{\cdot}50$ is about maximum. It is theoretically important to understand the causes of this apparent low correlation ceiling. But there is no doubt that the present evidence overwhelmingly rejects the null hypothesis. This is true of simple RT as well as choice RT (also termed discriminative or disjunctive RT). Both simple and choice RT are negatively correlated with IQ. The correlation between RT and IQ generally seems to be of the 'twisted pear' variety, i.e. the variation of RT around the linear regression of RT on IQ is not homescedastic. Comparing normals and borderline retardates, we find that whereas no normals have as slow RT as retardates' mean RT, a few retardates are as fast as normals. One might say that fast RT is necessary but not sufficient for a high IQ, whereas slow RT is sufficient but not necessary for a low IQ.

Mean differences in RT (or in various parameters of RT) between criterion groups selected for differences in ability as measured by sychometric tests or scholastic performance always give more clearly impressive evidence of a relationship between RT and general ability than the correlation coefficient. It may come as a surprise to many, as it did to me, that the mean RT difference between criterion groups is often of at least the same magnitude as the mean IQ difference between the groups, when the mean differences in RT and IQ are both expressed in standard deviation or σ units. We have found that borderline retarded young adults, with a mean IQ of about 70, differ from university students about 6σ on Raven's Matrices. These groups differ about 7σ (σ of the university students) in mean RT. University students compared with academically less highly selected students of the same age in a two-year vocational college differ about 1σ in scholastic aptitude scores; in mean RT they differ $1{\cdot}2\sigma$ in terms of the vocational college σ and $1{\cdot}9\sigma$ in terms of the university σ. So RT shows even quite substantial differences between groups of normal youths who differ in scholastic aptitude.

Mean differences of the order of only 50 – 100 ms are not at all noticeable to the naked eye while testing subjects on the RT apparatus. The layman is puzzled by the fact that such seemingly small differences in an absolute sense should be related to quite conspicuous differences in scholastic attainments, vocabulary, tests of reasoning ability, and the like. But if RT reflects some important aspect of information processing capacity, then even very small individual differences in *rates* of information processing, when multiplied by days, weeks, months, or years of interaction with the myriad opportunities for learning afforded by common experience, can result in easily noticeable differences in the amounts of acquired knowledge and developed intellectual skills. At a moment's glance there is scarcely a noticeable difference between the speed of a car averaging 50 and another averaging 51 miles per hour, but after a few hours on the road they are completely out of sight of one another.

From the standpoint of psychometrics, I think the most important conclusion from all the RT research is that it proves beyond reasonable doubt that our present standard tests of IQ measure, in part, some basic intrinsic aspect of mental ability and not merely individual differences in acquired specific knowledge, scholastic skills, and cultural background. The RT parameters derived from typical procedures cannot possibly measure knowledge, intellectual skills, or cultural background in any accepted meaning of these terms. Yet these RT parameters show significant correlations with scores on standard tests of mental ability and scholastic achievement and show considerable mean differences between criterion groups selected on such measures.

Three basic RT paradigms

There are three distinct and basic paradigms in RT research. Each paradigm measures different facets of information processing speed, and each has shown a relationship to psychometric variables. I shall refer to these paradigms by the names of the three psychologists who initiated them.

The *Hick* (1952) paradigm measures the linear increase in RT to visual or auditory stimuli as a function of the amount of information (measured as *bits*=$\log_2$ of the number of stimulus alternatives) conveyed by the reaction stimulus, but involves no need to access either short-term or long-term memory (STM or LTM). The classical experiment contrasting simple and two-choice RT is the simplest example of the Hick paradigm involving 0 and 1 bit of information, respectively.

The *Sternberg* (1966) paradigm presents the subject with a small set of digits (or letters), followed immediately by a single 'probe' digit to which the subject responds 'yes' or 'no' as to whether the probe was or was not included in the set. The *S*'s RT or decision time in pressing the 'yes' or 'no' key involves speed of scanning STM. and RT increases as a linear function of the number of *items* in the set. unlike the Hick phenomenon, in which RT increases as a linear function of the *logarithm* (to the base 2) of the number of stimulus alternatives. We still need a theoretical explanation of this fundamental difference between the Hick and Sternberg phenomena. It suggests that in making choices and decisions, the brain acts like a binary digital computer, whereas in STM scanning it behaves more like an analog device, paralleling the sensory system. STM scanning and visual scanning have been found to almost perfectly parallel one another, and individual differences in STM scanning and visual scanning are highly correlated (Chiang & Atkinson, 1976).

The *Posner* (1969) paradigm contrasts discriminative ('same' versus 'different') RTs to pairs of stimuli which are the same or different either physically or semantically. For example, the letters AA are physically the same whereas *Aa* are physically different but semantically the same. When *S*s are instructed to respond 'same' or 'different' to the physical stimulus, RTs are faster than when *S*s 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 LTM. Interestingly, Hunt (1976) and his co-workers have found that this measurement is especially related to verbal ability as measured by the Scholastic Aptitude Test (SAT-V) in university students.

The Hick, Sternberg, and Posner paradigms probably tap quite basic brain processes. The Hick and Sternberg phenomena are not peculiar to the human brain. Hick's law has been demonstrated in pigeons (Blough, 1977) and the Sternberg effect in monkeys (Eddy, 1973). Because the Posner effect involves semantic memory, no one has looked for it in infrahuman species.

Typical findings

Posner paradigm

Figure 1 shows the results of a study by Hunt (1976) using the Posner paradigm with groups of university students scoring high or low on the SAT-Verbal.

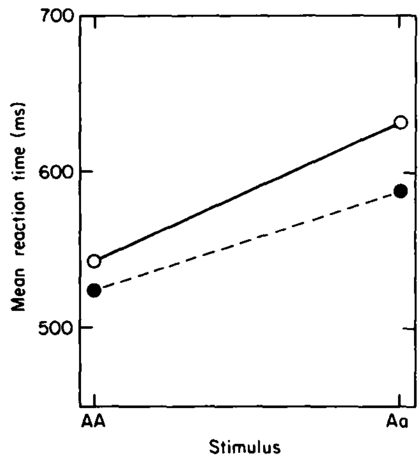

Fig. 1. Time required to recognize physical or semantic identity of letter pairs by university students who score in the upper (high) or lower (low) quartile on the SAT-Verbal (after Hunt, 1976, Table 1, p. 244) —o—, low verbal;- -●- -, high verbal

AA represents the physical identity choice (same-different) RT task; *Aa* represents the semantic identity task. University students require on the average about 75 ms more time to respond to *Aa* than to *AA* types, which is the time taken by semantic encoding of the stimulus. Two features of Fig. 1 are particularly interesting in relation to findings from the Sternberg and Hick paradigms: (1) the high and low groups on SAT-V show a mean difference in RTs even on physical, nonsemantic identity task, which is essentially just a form of classical two-choice discriminative RT; and (2) the mean RTs are all greater than 500 ms, which is appreciably slower than the RTs of university students in the Hick paradigm, even for RT to three bits (i.e. eight stimulus alternatives) of information, which has a mean RT of 350 – 400 ms. Because the times needed for physical discrimination between extremely familiar stimuli and for accessing simple, highly overlearned semantic codes in LTM are in excess of the RTs to three bits of information in the Hick paradigm, it suggests that performance in our Hick paradigm does not depend on discriminating anything as difficult as familiar letters or accessing anything in LTM. 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 our Hick paradigm with university students.

Sternberg paradigm

Figure 2 shows Sternberg STM-scan RTs for groups of fifth and sixth grade children with moderate and high IQs, from a study by McCauley *et al.* (1976).

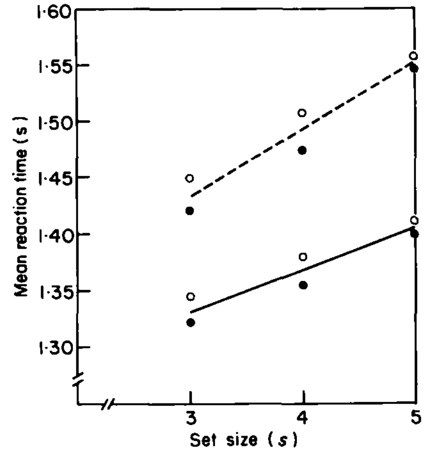

Fig. 2. Mean RTs for correct 'yes' and 'no' (i.e. presence or absence of probe digit in target set) for moderate IQ (95 or below, $\bar{X}$ = 88) and high IQ (115 or above, $\bar{X}$ = 126) fifth and sixth grade children. The equations for the two lines are: moderate IQ RT = 1265 + 58s, and high IQ RT = 1210 + 40s, where RT is in ms and s=number of digits in the target set, (From McCauley et al, 1976).----, moderate IQ; –, high IQ,○, no;●,yes

The intercepts and slopes of the moderate and high IQ groups both differ significantly. Stanford University students given a comparable Sternberg task (Chiang & Atkinson, 1976) show much lower intercepts (about 400 ms) but show about the same slope (i.e. scan rate of 42 ms per digit in target set) as the high IQ children (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 has 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, Marshalek & Lohman (1976) were able to 'predict' the intercepts and slopes of the Sternberg memory scan paradigm for individual Stanford students with multiple R's of 0·88 and 0·70 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 0·33 and 0·56. SAT-Verbal and SAT-Quantitative scores were predicted with R's of 0·54 and 0·21, respectively. Remember, we are dealing here with the quite restricted range of ability in Stanford University students.

Hick paradigm

This is an elaboration of simple and choice RT. Hick (1952) discovered that RT increases linearly as a function of $\log_2$ of the number of choices or stimulus

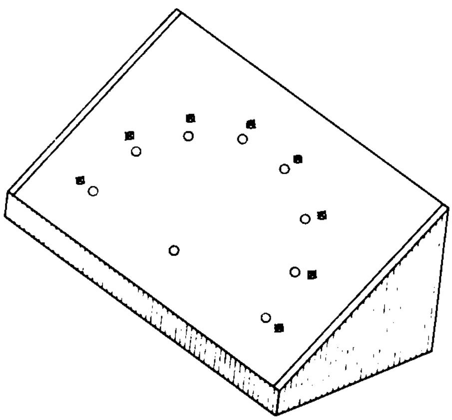

Fig. 3. Subject's console of the reaction time-movement time apparatus. Push buttons indicated by circles, green jeweled lights by circled crosses. The 'home' button is in the lower center

alternatives – a phenomenon now known as Hick's Law. I have been doing studies of this paradigm, using an apparatus shown in Fig. 3. (It is described in more detail by Jensen & Munro, 1979.) The S places his index finger on the "home" button, a "beep" warning signal is sounded for 1 s, and after a random interval of 1 - 4 s one of the green lights goes on. The S must turn off the light as fast as possible by touching the button adjacent to it. The time between the light's going on and removal of the S's finger from the home button is the RT. The interval from release of the home button to turning out the light is the movement time (MT). Templates can be placed over the console to expose any number of light/button alternatives from 1 - 8. We have most often used

1,2,4 and 8 alternatives, corresponding to 0,1,2 and 3 bits of information. Following instructions and several practice trials, *S*s are usually given 15 trials on each number of alternatives (60 trials in all) in a single session lasting about 20 minutes. Roth (1964) was the first to find a correlation between the slope of RT (confounded with MT in his study) as a function of bits and mental age.

The cognitive demands of this task are so extremely simple that it seems almost implausible that the procedure could yield any measurements that would be correlated with IQ. Even normal three-year old children and institutionalized retarded adults with IQs below 20 can meet the simple task requirements.

Therefore, the initial aim of my research with this paradigm has been to establish that the parameters measured by the procedure are in fact correlated with intelligence. If there is a correlation, then we can go to more refined process or componential analysis of the paradigm with a view to developing a theory to explain the observed phenomena. In so doing, we should discover something fundamental about the nature of intelligence, that is, Spearman's *g* which to this day remains an unconquered frontier of psychology (Jensen, 1979).

To insure that the RT phenomena are in fact related to intelligence, I have sought correlations between RT parameters and IQ in criterion groups selected from every available level of the IQ distribution, ranging from the severely retarded (with IQs of 15 - 50), to the mildly retarded and borderline (IQs 50 - 80 or so), to average and bright school children and average young adults, and to university students with IQs above the 95th percentile of population norms. We have now tested nine such groups totalling about 800 persons. Without exception, groups differing in mean IQ also differ very significantly in the expected direction in a number of RT (and also MT) parameters. Also *within* every group we have tested, the RT parameters are significantly correlated with IQ, with all correlations in the theoretically expected direction, mostly ranging between about 0·20 and 0·50.

Figure 4 shows mean RT and MT of 39 ninth-grade girls grouped into low, middle, and high thirds of the distribution of scores on Raven's Standard

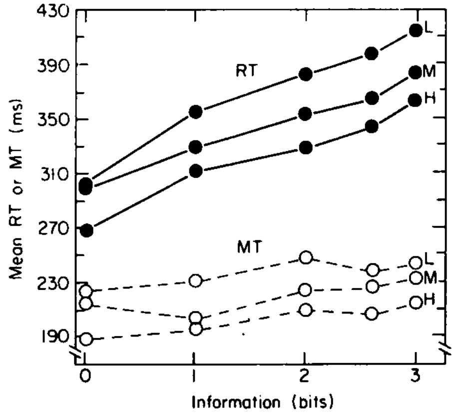

Fig. 4. Mean RT and MT for the high (H), middle (M), and low (L) thirds of a sample of ninth-grade girls on Raven's Standard Progressive Matrices (from Jensen & Munro, 1979)

Progressive Matrices. RT increases linearly with bits, whereas MT shows no appreciable change over increasing information. Yet RT and MT are both correlated with Raven scores. The multiple R of RT and MT with Raven is 0·50. Figure 5 shows RT and MT of 50 university students; the vertical lines indicate the average intraindividual variability (the mean of the standard

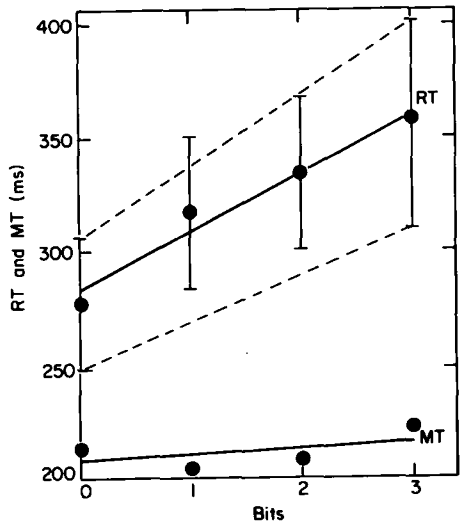

Fig. 5. Mean median RT and MT, and the mean standard deviation of RT (vertical lines) of 50 university students

deviations over 15 trials for each individual). Note the disparity between RT and MT. We find that disparity between RT and MT is related to the average intelligence level of our various criterion groups, with university students showing much slower RT than MT and the most retarded group (mean IQ of 39) showing the reverse, i.e. faster RT than MT. A plot of the RT/MT ratio i.e. the ratio of the mean simple RT (i.e. 0 bits) to the mean MT, as a function of the average intelligence levels of our four adult criterion groups is shown in Fig. 6. Individual differences in RT and MT are not highly correlated and

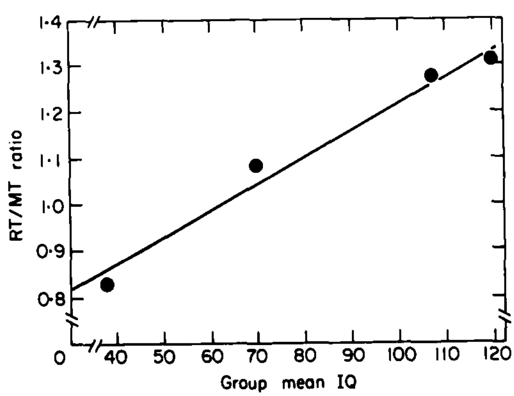

Fig. 6. Ratio of mean of simple RT (0 bits) to mean MT plotted as a function of average intelligence levels of adult criterion groups: severly retarded, borderline retarded, junior college students, and university students

apparently reflect different processes. RT shows markedly less day-to-day test-retest stability than MT, and MT is not significantly correlated with intelligence in our university students, although it is significantly correlated with IQ in children and retarded adults. Figure 7 shows RT and MT of 46 boderline retarded young adults (mean IQ of 70) grouped above and below

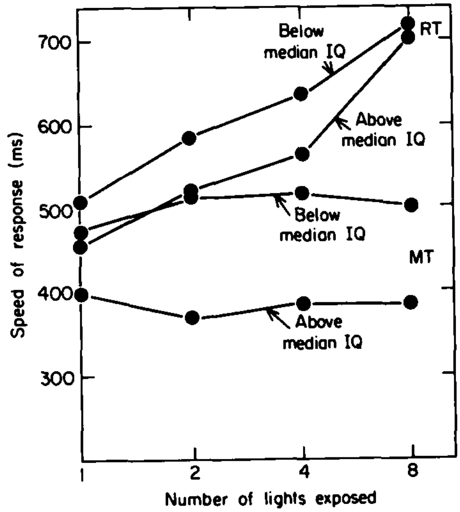

Fig. 7. Mean RT and MT of 46 borderline retarded young adults grouped above and below median on Raven's Standard Progressive Matrices (from Vernon, 1979)

the median on Raven SPM scores. MT shows a larger difference between the upper and lower halves of the sample in Raven scores than does RT.

RT parameters and intelligence

We describe an individual's RT performance in the Hick paradigm in terms of three parameters: the *slope* of the linear regression of RT on bits, the *intercept* of the regression line, and the *intraindividual variability* over trials, which is indexed by the root mean square of the variances among trials *within* bits. (We have also used the slope of the regression of the standard deviation among trials, as a function of bits.) The parameters are shown in Fig. 8 from averaged data on 46 borderline retarded young adults. There is zero slope for MT, which is typical. RT increases according to Hick's law, and the intraindividual variability increases exponentially across bits. The slope of the variability is more highly correlated with psychometric *g* than any other RT parameter. There are reliable individual differences in all of the RT parameters I have mentioned, and all of them are positively intercorrelated. Other investigators, too, have found a positive correlation between intercepts and slopes in the Sternberg paradigm (Dugas & Kellas, 1974; Snow *et al.*, 1976). Oswald (1971) found an *r* of +0·46 between simple RT (intercept) and the slope of RT over 15 card-sorting tasks of graded complexity. Moreover, all these parameters are negatively correlated with *g*. At first I expected that intercepts; which represent simple RT, and hence involve little or no information processing, would not be correlated with IQ. I was wrong;

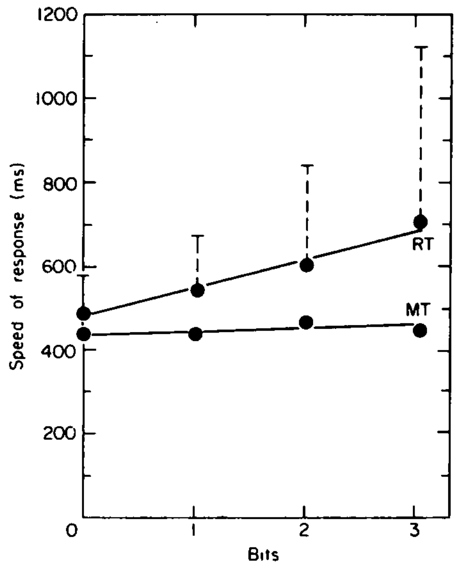

Fig. 8. Mean RT and MT, as a function of bits, and mean intraindividual variability (vertical dashed lines = intra-individual standard deviation of RT over 15 trials), in 46 borderline retarded young adults (from Vernon, 1979)

intercepts are negatively correlated with IQ, although within fairly homogenous criterion groups the correlations are often too small to be significant and are almost invariably smaller than the correlations of slope and intraindividual variability with IQ. Figure 9 shows the intercepts and slopes of RT data from

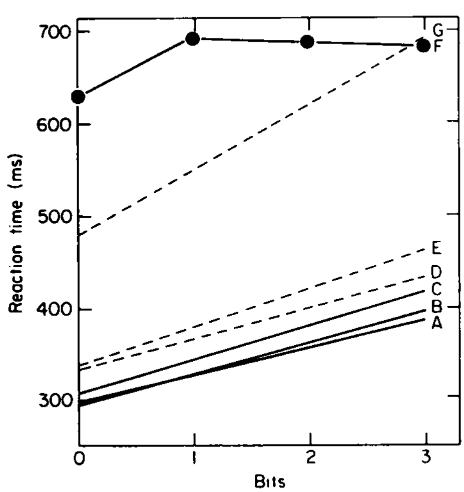

Fig. 9. RT as a function of bits, illustrating Hick's law and differences in intercepts and slopes, for diverse groups varying in age and intelligence: A, university students; B, ninth grade girls; C, sixth graders in a high SES-high IQ school; D and E, white and black, respectively, male vocational college freshmen with approximately equal scholastic aptitude scores; F, severly mentally retarded young adults (mean IQ 39); G, mildly retarded and borderline young adults (mean IQ 70) (from Jensen, 1979)

six criterion groups. None of the regression lines except that of the severely retarded group shows a significant nonlinear trend.

Intercept

The intercept is the most complex of the parameters and is most susceptible to variations in experimental procedure and apparatus. Even with such different procedures as the Hick and Sternberg paradigms, the slopes are remarkably similar (30 - 40 ms per bit or item) in groups of comparible age and intelligence, while the Sternberg intercept is about 100 milliseconds greater than the Hick intercept. We have found that the RT intercept is reduced about 30 ms in university students by requiring *S*s simply to release the home button as fast as possible, without having to make the ballistic response to turn out the light. (The slope is unaffected by this procedure.) The additional RT time required by having to make a precise response following release of the home button is probably an example of Fitt's law, which states that the delay (RT) in response following a signal to respond is a monotonically increasing function of the complexity or precision required of the response; i.e. some small part of the *S*'s RT consists of preprogramming the response that follows the reaction stimulus, so that a more precise movement, as is required for turning out the light, results in a slightly longer RT than the RT for merely having to get off the home button without any further precise response. It remains to be determined how this procedural variation will affect the correlation between RT intercept and IQ.

The intercept is also affected by sensory modality, being higher for visual than for auditory stimuli. The intercept of RT includes such physiological processes as sensory lag and speed of peripheral nerve conduction. We clearly need a thorough compential analysis of the RT intercept to determine which components are responsible for the correlation of the RT intercept with IQ. At this point I will not be too surprised if we find IQ-related individual differences in sensory lag and speed of nerve conduction. There are likely individual differences in such basic processes. Hegmann (1975) found differences in speed of peripheral nerve conduction in the caudal (tail) nerves of mice and was able, with five generations of selective breeding, to obtain two strains of mice that differed by more than 20 per cent in speed of peripheral nerve conduction. Although selection was based on speed of conduction in the caudal nerve, selective breeding produced a generalized effect on the other peripheral nerves, and the slow and fast neural conduction strains showed behavioral differences at the reflex level.

Intraindividual variability

Surprisingly little attention was ever given to intraindividual variablity in RT in the older literature. Yet it is this aspect of individual differences in RT that seems to be the most profoundly related to intelligence level, as has been frequently noted by investigators of RT in the mentally retarded (Berkson & Baumeister, 1967; Baumeister & Kellas, 1968*a*, 1968*b*, 1968*c*; Liebert & Baumeister, 1973; Wade, Newell, & Wallace, 1978; Vernon, 1979). The negative correlation between intraindividual variability in RT and IQ is found within every level of intelligence, from the severely retarded to university students. Intraindividual variability, henceforth labelled σ_i, is measured by the standard deviation of an individual's RTs over repeated trials obtained in a single session. Even university students and vocational college students (all whites) show a highly significant (t=5·23,p<0·001) difference (0·68σ) in mean σ_i for simple RT. A theory of the relationship between RT and IQ will have to explain σ_i, which may actually be an even more basic phenomenon than the relationship of mean or median RTs to IQ. The median and especially the mean RT are not independent of σ_i. Figure 10 shows the typical highly skewed frequency distributions of the simple RTs of six retarded (IQs 50 - 81, mean IQ 62) but physically normal persons and six university students of about

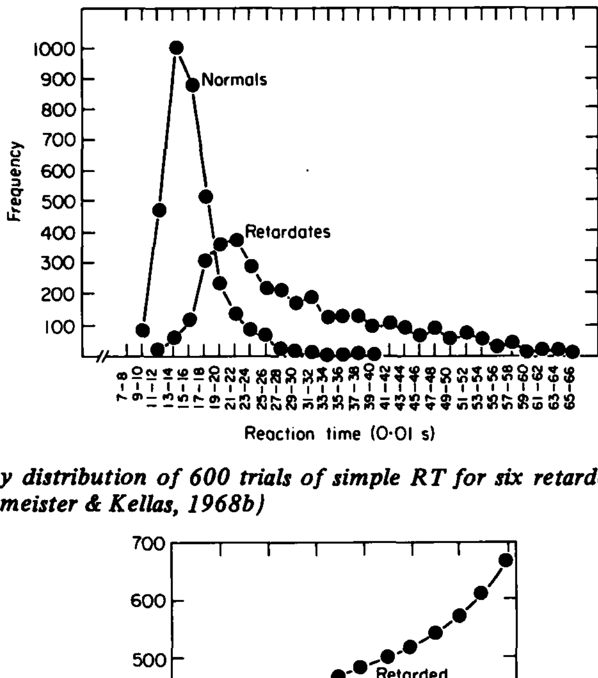

Fig. 10. Frequency distribution of 600 trials of simple RT for six retarded and six normal persons (from Baumeister & Kellas, 1968b)

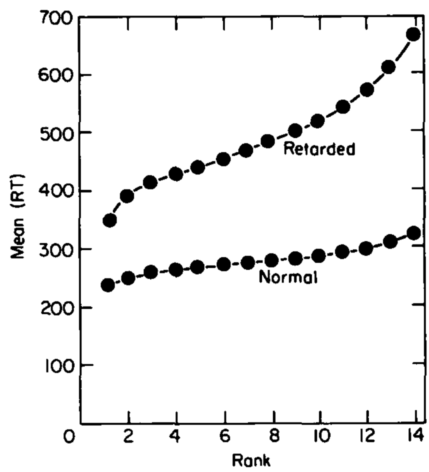

Fig. 11. Mean simple RT plotted after ranking RTs on 15 trials from the fastest to the slowest trial (omitting the 15th rank) for retarded and normal Ss

the same age, each given 600 trials. I think that the most interesting feature of Fig. 10 is that in a total of 3600 trials of *simple* RT, the retardates do not produce any RTs that are as fast as the fastest RTs of the normal *S*s. This seems to me to be an extremely important phenomenon that will figure prominantly in future theoretical formulations. Liebert & Baumeister (1973) have found high correlations (as high as 0·96 in college students) between mean RT over 100 trials and the average of the ten shortest RTs in 100 trials. They also note that the σ_i decreases with age between the ages of about 6 – 18 years, and the lower limit of RT decreases with age in that age-range.

I have looked more closely at this phenomenon in our data by rank ordering each *S*'s RTs from the shortest to the longest in 15 trials. (The 15th rank is eliminated to get rid of possible outliers.) Figure 11 shows the means of the ranked RTs of 46 mildly retarded (IQ 70) and 50 bright normal (IQ 120) young adults each given 15 trials on simple RT. Note that even on the fastest trial (rank 1) the retarded and normal *S*s differ by 111 ms. In fact, the normal *S*s' *slowest* RT (rank 14) is 32 ms shorter than the retardates' *fastest* RT. The difference becomes more exaggerated for choice RT for three bits (i.e. eight light/button alternatives), as shown in Fig. 12. The fastest RTs of the normals

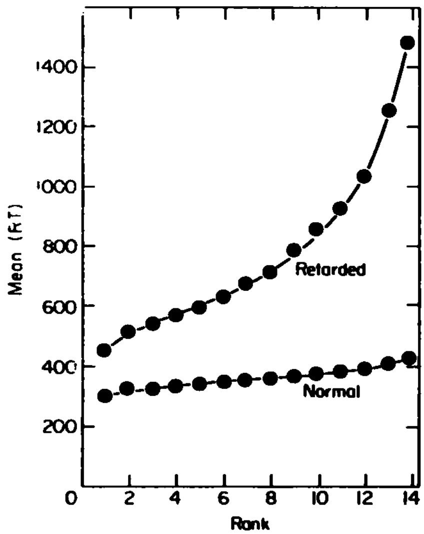

Fig 12. Mean choice RT for 3 bits in Hick paradigm, plotted after ranking individual RTs on 15 trials from the fastest to the slowest trial (omitting the 15th rank) for retarded and normal Ss

and retardates here differ 142 ms. [A logarithmic transformation (i.e. $\log_{10}$ RT) does not essentially alter this picture.] In case anyone might think these are trivial differences, let us look at them in terms of standard deviation or σ units, i.e. (normal RT minus retarded RT)/σ, as shown for simple RT in Fig. 13 for σ differences based on both normal and retarded σ units. The fastest simple RT of retardates and normals differs 1·2σ in terms of the retardates' σ units and 4·8σ in terms of the normals' σ units. Remember, Fig. 13 is *simple* RT.

The fact that even the fastest RTs of the retarded *S*s are slower than the RTs of normals, even for simple RT, suggests that the difference is at some very basic, one might almost say neural, level and not at any very complex level of information processing. Possibly even simpler responses might show reliable speed differences related to general intelligence.

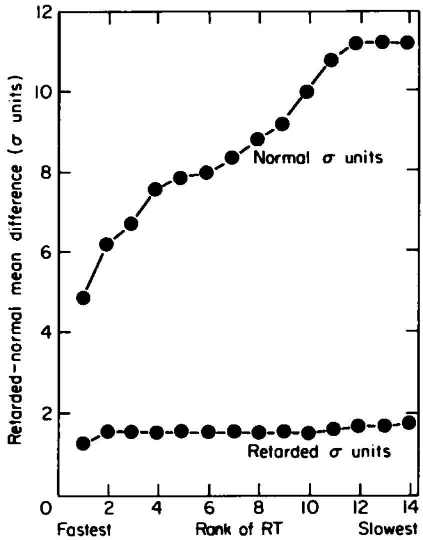

Fig. 13. Difference in simple RT between retarded and normal Ss, expressed in both normal and retardate σ units, with RTs for 15 trials ranked from fastest to slowest

Combining RTs in the Hick, Sternberg, and Posner paradigms

If RT and the derived parameters in the three different paradigms reflect different processes, involving stimulus encoding, scanning of STM, and retrieval of semantic codes in LTM, all of which are probably involved in arriving at the correct answers to the relatively complex items used in ordinary intelligence tests, we should expect that an optimally weighted combination of RT measurements derived from all three 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 & Bobbitt (1978) found. Three RT-derived measures were obtained on each *S*: (1) choice RT minus simple RT (Hick paradigm), (2) semantic minus physical same/difference 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 in grades 3,7 and 11 was 0·59, 0·57, and 0·60, in the three grades, respectively. I imagine that still higher correlations would 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 0·27, indicating that they are tapping different processes as well as sharing some variance in common. A most interesting finding of this study is that inferred similar processes across the three paradigms show higher intercorrelations (average r=0·66) than inferred dissimilar processes (average r =0·30). So there appears to be a general factor plus specific factors for the processes inferrable from the three paradigms and derived from the basic RT measurements.

The burning question is this: Will it be possible to discover a small number of such basic processes, measurable by means of RT, that will yield parameters which, in an optimally weighted combination, will 'account for' practically all of the true *g* variance in psychometric tests of mental ability? Might not differently weighted combinations of a few process measurements based on RT also account for the variance in the so-called group factors involved in verbal, quantitative, and spatial abilities? This is what we must try to find out. Whatever the outcome may be, the effort will be amply rewarded by the gain in our theoretical understanding of the nature of mental abilities, to say nothing of the potential for practical applications should it turn out that most of the variance in complex mental abilities now measured by psychometric tests can be accounted for in terms of a number of RT parameters in a few fundamental paradigms.

Task complexity and the RT-intelligence correlation

Simple RT correlates less with IQ than does choice RT; in general, in the Hick paradigm, as the bits of information conveyed by the response stimulus increase, the higher is the correlation between RT and IQ (Jensen, 1979; Jensen & Munro, 1979; Lally & Nettlebeck, 1977). A similar generalization holds for the Posner paradigm (Hunt, 1976; Goldberg, Schwartz & Stewart, 1977). But 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 be extend beyond the range of tasks to which RT is greater than about 1000 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 & 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 conditions, 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 non-significant correlations are between individual differences in response times to test items and total scores (i.e. number right) on the test. When solution times for items are *averaged* over *S*s, the correlation between mean item solution times and item difficulties (i.e. proportion of *S*s attempting the item but failing to get the right answer) approaches unity (Elliott & 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 *S*s should be compared on only those items that all *S*s get right, otherwise the response times of the brighter *S*s 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 correlation between test scores and RT parameters derived from relatively simple paradigms in the 0 to 3 bits range of information processing demands. Figure 14 illustrates hypothetically what I suspect happens as the information processing demands of the task increases.

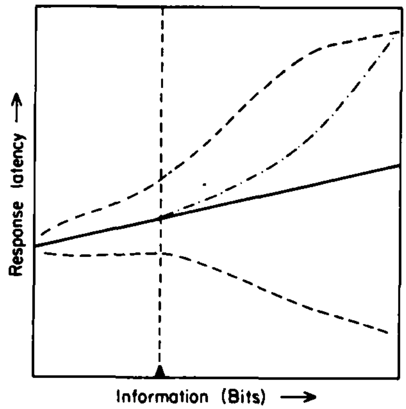

Fig. 14. Hypothetical illustration of increase in response latency (solid line) as a function of information, with an increasing spread of individual differences (dashed line) around the mean response latency, and a departure of linearity (dot-dash-dot line) of response latency beyond a certain information load (vertical dashed line)

We know that both intra- and inter-individual differences in RTs increase with increasing amounts of information in the response stimulus. Beyond a certain point, however, the correlation scatterdiagram bulges out in a way that prevents a high correlation, even though the mean RT over *S*s continues to be a monotonically increasing function of the amount of information. Moreover, the *nominal* information in the stimulus 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 STM while performing other opertations. So with increasing task complexity, beyond a certain point, the RT departs from linearity, rising at a positively accelerated rate. This is nicely illustrated in Fig. 15 from a study by Royer (1977) in which Kohs block designs were scaled in terms of bits according to the number of binary possibilites of the positions of the blocks in a given design. One element of the information processing demands was minimized or eliminated in the cued condition by heavy black lines on the design target cards indicating the dividing lines between the blocks making up the pattern.

Such considerations underline the importance of componential analyses of complex tasks, in which RTs are determined for each component, and these can then be combined in a multiple regression equation to yield a multiple *R* with psychometric measures of ability. The importance of an optimally weighted sum (i.e. multiple regression) of the times for the various

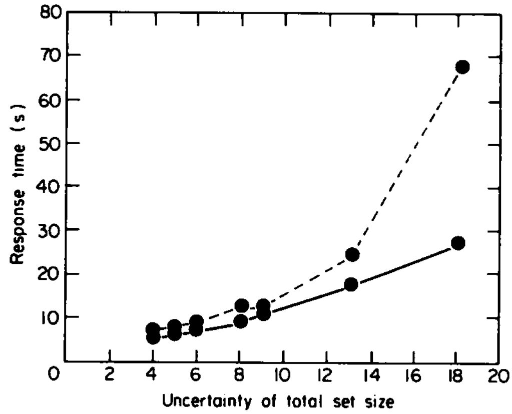

Fig. 15. Mean response times for cued and uncued stimuli in the Kohs block design test as a function of uncertainty or amount of information (i.e. binary possibilities) in the block design (from Royer, 1977). ●–, *Cued;* ——, *uncued*

component processes, rather than the total time (i.e. the sum of the unit-weighted component times) becomes obvious when we realize that the time for some processes may be *positively* correlated with performance on a complex task while the time for other processes may be *negatively* correlated with performance. Robert Sternberg (1977) has found that high IQ persons take *longer* to encode analogy items than do less intelligent persons. But more thorough stimulus encoding, which takes more time, leads to more efficient solutions at later stages of processing.

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 1000 ms among university students. Yet total time on Raven's matrices was found to be correlated – 0·46 with E (extraversion) scores on the Eysenck Personality Inventory, whereas the correlation between total time and Raven scores was exactly zero.

RT in relation to other variables

In this review of empirical findings I have not attempted a theoretical formulation of the relationship between RT and intelligence. The explication of a theory calls for more detailed descriptions of empirical phenomena than I can present in this brief paper. Important methodological aspects of RT research involving questions of the optimal measures of various parameters and the reliability and temporal stability of measurements also need to be considered. The relationship of RT to more elemental physiological variables, such as average evoked potentials, metacontrast (visual masking), critical flicker frequency, and the effects of various drugs on RT, are essential parts of the theoretical picture. Developmental trends in RT, on which there is already a considerable literature, are also grist for theory. There seem to be sex differences in RT phenomena, too, but their nature and cause are still obscure (e.g. Chiang & Atkinson, 1976). RT in the Posner paradigm is also related to specific reading disability (Spring, 1971), which suggests the possibility of using various RT measurements for the diagnosis of specific educational handicaps. In a more elaborate forthcoming paper (Jensen, in preparation), I will indicate the relevance of evidence from all these lines of RT research to a general theory of intelligence.

References

Baumeister, A. A. & Kellas, G. (1968*a*). In *International Review of Research in Mental Retardation*, Vol. 3. (N. R. Ellis, Ed.), New York: Academic Press.

Baumeister, A. A. & Kellas, G. (1968*b*). *Am. J. ment. Defic.* 72, 715.

Baumeister, A. A. & Kellas, G. (1968*c*). *J. Abnor Psychol.* 73, 421.

Berkson, G. & Baumeister, A. A. (1967). *Am. J. ment Defic.* 72, 262.

Blough, D. S. (1977). *Science* **196**, 1013.

Chiang, A. & Atkinson, R. C. (1976). *Mem. Cognit.* **4**, 661.

Dugas, J. & Kellas G. (1974). *J. Exp. Child Psychol.* **17**, 177.

Eddy, D. R. (1973). Memory processes in *Macaca* speciosa: Mental processes revealed by reaction time experiments. Unpublished doctoral dissertation, Carnegie-Mellon University.

Elliott, C. D. & Murray, D. J. (1977). *Br. J. educ. Psychol.* **47**, 50

Gilbert, J. A. (1894). *Stud. Yale Psychol. Lab.* 2, 40.

Goldberg, R. A., Schwartz, S. & Stewart; M. (1977). *J. educ. Psychol.* **69**, 9.

Hegmann, J. P. (1975). *Behavl. Biol.* **13**, 413.

Hick, W. (1952). *Q. J. exp. Psychol,* **4**, 11.

Hunt, E. (1976). In *The Nature of Intelligence.* (L. B. Resnick, Ed.), p. 237–259. Hillsdale, N.J.: Erlbaum.

Jensen, A. R. (1979). *Creative Sci. Technol.* 2, 16.

Jensen, A. R. In *Advances in the Psychology of Human Intelligence*, (R. J. Sternberg, Ed.) in preparation.

Jensen, A. R. & Munro E. (1979). *Intelligence* 3, 121.

Keating, D. P. & Bobbitt, B. (1978). *Child Devel.* **49**, 155.

Lally, M. & Nettelbeck, T. (1977). *Am. J. ment. Defic.* **82**, 273.

Liebert, A. M. & Baumeister, A. A. (1973). *J. Psychol.* **83**, 57.

McCauley, C., Dugas, J. Kellas, G. & DeVellis, R.F. (1976). *J. educ. Psychol.* **68**, 474.

Oswald, W. D. (1971). *Psychologische Rundschau* **22**, 197.

Posner, M. I. (1969). In *The Psychology of Learning and Motivation* pp. 43-100. (Vol. 3). (G. H. Bower & J. T. Spence, Eds) New York: Academic Press.

Roth, E. (1964). *Z. Exp. Angew. Psych.* **11**, 616, 622.

Royer, F. L. (1977). *Intelligence* **1**, 32.

Snow, R. E. Marshalek, B. & Lohman, D. F. (1976). *Technical Report No. 3* Aptitude Research Project, School of Education, Stanford University.

Spiegel, M. R. & Bryant, N. D. (1978). *J. educ. Psychol.* **70**, 904.

Spring, C. (1971). *J. Educ. Psychol.* **62**, 492

Sternberg, R. J. (1979). *ONR Technical Report, No. 17* Department of Psychology, Yale University, New Haven, Conn.

Sternberg, S. (1966). *Science* **153**, 652

Vernon, P. A. (1979). Reaction time and intelligence in the mentally retarded (unpublished paper).

Wade, M. G., Newell, K. M. & Wallace, S. A. (1978). *Am. J. ment. Defic.* **83**, 135.

White, P. O. (1973). In *The Measurement of Intelligence.* (H. J. Eysenck, Ed) pp. 246–260. Baltimore: Williams & Wilkins.

Appendix

Glossary of reaction-time terminology

Bit. From "*b*inary dig*it*," a term used in information theory and measurement to express a *unit* of information. A *bit* is defined as the amount of information required to reduce uncertainty by one-half, or as the equivalent of the result of a binary choice, i.e. a choice between two alternatives (as "yes" or "no," or "on" or "off"). In RT experiments, the amount of information in the stimulus display, measured in *bits*, is the logarithm (to the base 2) of the number of stimulus alternatives.

Decision time. Same as RT, but sometimes used instead of RT to emphasize the distinction between the speed of the observer's first overt response to the onset of a stimulus (such as releasing a depressed telegraph key), which is termed *decision time*, and making a second overt response (such as pressing the same key or another key to turn off the stimulus), which is termed *movement time.*

g. The *general factor* of mental ability or intelligence. The *g* factor is estimated by means of a variety of mathematical techniques generally referred to as *factor analysis*, which is based on the matrix of all the intercorrelations among a large number of diverse mental tests. The *g* factor reflects the variance (individual differences) that is measured in common by all of the tests. The extent to which any particular test measures the *g* factor that is common to all of the tests is termed the tests' *g loading*. The square of a particular tests' *g* loading is the proportion of variance in scores on that test which is common to all of the other tests that entered into the factor analysis.

Intraindividual variability. The individual's variability in RT (or MT) from trial to trial, when a number test trials are given, as indexed by the standard deviation of the individual's RT measurements taken over the number of trials. (Cf. Reaction time parameters.)

LTM. Long-term memory; a memory trace usually of longer than one-minute's duration. (Cf. *STM.*)

Movement Time (MT). The time interval *between* the observer's first overt response to a stimulus and a second response the observer may be required to make – for example, releasing one telegraph key at the onset of the stimulus and then as quickly as possible pressing another key. The time interval between the observer's release of the first key and the response to the second key is termed *movement time.*

Psychometric abilities. Abilities that are measured by means of psychometric tests, that is, tests consisting of a number of items of graded difficulty (i.e. questions, problems or tasks) on each of which the person's performance is scored as passed or failed, so as to yield a total raw score indicating the number of items passed.

Raven's matrices: also *progressive matrices.* A highly *g*-loaded non-verbal test of reasoning ability consisting of figural patterns and geometric forms which are universally found in all cultures.

Reaction Time (RT). A general term referring to the interval of time (usually measured in milliseconds) between the onset of a stimulus and the beginning of the observer's overt response. The character of the response is prearranged through experimental conditions and instructions given to the observer.

There are several sub-types of RT:

Simple RT. The observer's RT to the onset of a single stimulus, e.g. a light or a tone.

Choice RT. The observer's RT to one stimulus out of a set of two (or more) stimulus alternatives, any one of which could occur at random; e.g. pressing (or releasing) a telegraph key when a red light comes on, but not when a green light comes on.

Discriminative RT. Generally, the same as *choice* RT, i.e. on overt response to one of two or more stimuli that may be presented, although it may also involve keenness of sensory discrimination when there are only light differences between the stimulus alternatives to which the observer responds.

Conjunctive RT. The observer's RT to the onset of two (or more) stimuli occurring *simultaneously* and the withholding of response to any stimulus presented singly.

Disjunctive RT. The observer's RT to the onset of *either one* of two (or more) stimulus alternatives out of a larger set of alternatives; e.g. responding to a red *or* green light, but withholding response to white, yellow and blue lights.

Reaction-time parameters. Any of the various statistical features that can be derived from the RT data obtained from a single observer on whom a large number of RT measurements have been obtained. The most commonly used parameters are (1) the *mean* RT of the observer measured in a number of test trials, (2) the *standard deviation* of the observer's RT measurements over a number of test trials, which is an index of *intra-individual variability*, and (3) in experiments involving simple and choice RT of varying levels of complexity, the *intercept* and *slope* of the linear regression of the observer's RT as a function of the levels of complexity. (Complexity is often measured in terms of *bits* of information.)

Response latency. Essentially the same as reaction time, but more often used when the RT experiment involves a highly complex reaction stimulus (and therefore the RT is relatively slow), such as an item from a standard intelligence test; e.g. the time interval between the presentation of a test item to a person and the person's stating (or marking) his answer.

STM. Short-term memory; a memory trace usually of less than one-minutes duration, or a test for the memory of an event that had occurred within the immediate past of one minute or less, regardless of how long the event may be remembered thereafter. (Cf, *LTM*.)